Interoception, emotional arousal and harm-based moral dilemmas (Alternative Format Thesis)

Brown, Helen

Award date:
2021

Awarding institution:
University of Bath

Interoception, emotional arousal and harm-based moral dilemmas

Helen Rhian Brown

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Psychology

November 2020

Acknowledgments

I would like to express my sincere gratitude to my PhD supervisors Prof. Danaë Stanton Fraser and Dr Michael Proulx for granting me the opportunity to carry out a PhD in an exciting and innovative Psychology Department at the University of Bath, and for their consistent support, encouragement and belief in me during these past three years. Their guidance, insightful questions and exposure to a variety of research projects and opportunities was instrumental to shaping my PhD and has allowed me to develop my research skills in ways I did not expect. I am also extremely grateful for their generosity in time and attention during more difficult points of this journey, and their personable and caring approach to supervision which is felt by their students and contributes to strong sense of support and team-spirit within the labs. I would like to sincerely thank all members from the CREATE Lab and Crossmodal Cognition Lab for their friendship, support and inspiration, which has helped me to develop as a researcher and deeply enriched my PhD experience. I would also like to thank Dr Susanna Martin, Dr Mary Nikolaidou and Ms Caroline Ransford for their technical and administrative support which has been critical to my PhD research but also for their key role in creating a positive and welcoming experience in the department. I would also like to thank Dr Harry Farmer, Dr Chris Bevan, Dr David Green, Mrs Jo Gildersleve, Prof. Ki Cater and Prof. Mandy Rose, from the EPSRC funded VR Documentary Encounters project for their support, wisdom, and sense of humour during my time on the project. I am very grateful to have been a part of this exciting project and to have had the opportunity to learn from researchers with such an interesting range of experience. I would also like to thank Dr Chris Ashwin and Dr Punit Shah, for the interesting discussions and thoughtful questions during my PhD confirmation viva which helped me focus my research at a formative point. I would like to thank Dr Kathryn Francis, for her guidance and interesting discussions about conducting research using virtual reality moral dilemmas, and my Dad, Steve, and Dr Marc Holmes for their technical support and relentless problem-solving attitude when building the virtual reality environments. Importantly, I would like to sincerely thank all of the research participants who gave up their time to contribute to this research, without whom this PhD would not have been possible. Finally, I would like to thank my family and friends for their unending support and patience. I would particularly like to thank my partner Alex, who has been my anchor, cheerleader and companion through this extraordinary journey. I would also like to thank my Mum, Lyn, Dad, Steve, and sisters Julia and Imogen for their encouragement, support and unwavering belief in me which has been a consistent source of motivation and lifted me up when I needed it most.

Copyright notice

Restrictions on use and Licensing

Access to this thesis/portfolio in print or electronically is restricted until …….. *(date).*

Signed on behalf of the Doctoral College (print name) ……..

Declarations

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Candidate's signature

This information has been redacted for privacy reasons.

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally.

Candidate's signature

This information has been redacted for privacy reasons.

Table of Contents

Abstract

Moral decisions are some of the most important decisions we make; often resulting in significant consequences for ourselves and others. In circumstances where there is little time for deliberation (Kahneman, 2003) or we are uncertain about the right course of action, we are likely to rely on other forms of emotional and bodily intuition to guide our decision-making or behaviour (Damasio, 1996). There is a wealth of research supporting the relevance of somatic signals for guiding moral decision-making. In particular, in response to moral dilemmas concerning the harm of others, a negative arousal response pattern is associated with inclinations to reject harmful action (e.g. Cushman et al., 2012; Greene et al., 2001). Importantly, the extent to which we perceive and interpret internal sensations varies a great deal between individuals. Interoception is a multi-dimensional concept (Garfinkel & Critchley, 2013) which refers to our perceptual experiences of visceral sensations. There has been no prior empirical research specifically investigating individual differences in interoception in the relationship between emotional arousal and moral decision-making. This PhD thesis explored whether individual differences across a range of interoceptive dimensions influenced moral judgments and actions, and whether this was linked to concurrent visceral events and experiences such as physiological arousal, emotional state or hunger. Cardiac and gastrointestinal measures of interoception were used to understand whether interoceptive sensitivities across these systems were aligned or diverged in their relationship with moral judgment and arousal. Text-based and immersive virtual reality moral dilemmas were used to explore moral judgments and behaviour. We found individual differences in interoceptive sensibility indirectly predicted harm aversion responses to egocentric moral judgments. In contrast, subjective sensations of hunger predicted allocentric judgments of unprofitable harmful acts. Most significantly, we found an ability to consciously direct attention to heartbeats modulated the relationship between changes in physiological arousal and moral judgments and behaviour. Finally, we found that a tendency to worry about painful or unpleasant bodily sensations may predispose people to appraise the harm and suffering of others depicted in coronavirus media articles as more salient; appraisals which predicted socially relevant behavioural intentions during the pandemic.

Introduction

When facing difficult situations which threaten our personal or moral values we are often advised to *"Trust your gut*" or *"Listen to your heart".* These phrases carry an implicit assumption that subtle and often impalpable physiological sensations can provide value in guiding us towards the 'right' decision. Interoception is a perceptual capacity that refers to our experiences of such visceral sensations inside the body which are inherently associated with processes of homeostatic regulation and error-perception (Craig, 2015, Ullsperger et al., 2010). Interoception is a multi-dimensional construct (Garfinkel & Critchley, 2013) and can be explored across a range of visceral domains including the cardiovascular system, respiratory system, gastric sensitivity, and pain (Craig, 2015). There is growing evidence that our subconscious and conscious perception of visceral sensations shares a common neural basis with emotion, motivation and overt behaviour (Feldman Barrett et al., 2004; Craig, 2015; Damasio, 1996; Farb et al., 2015). Furthermore, there is increasing evidence for the central role of somatic markers in moral decision-making and moral action (Damasio, 1996; Greene et al., 2001; Teper et al., 2015; McDonald et al., 2017).

Yet, there is a lack of research investigating whether distinct facets of interoception, such as our interpretation of internal sensations or accuracy in perceiving them, are related to decision-making and behaviour in moral dilemma scenarios. Virtual-reality (VR) technology provides exciting opportunities for the study of moral behaviour, moving beyond traditional text-based thought experiments into the realm of immersive moral dilemmas with increasing sensorimotor control (Francis et al., 2017; Pan & Slater, 2011). Moreover, studying moral behaviour in more ecologically valid environments could provide useful applications for understanding 'real world' behaviour (Pan & Slater, 2011) such as in occupational contexts like the NHS or police force, where moral decisions are frequently made and have profound consequences for the wellbeing of others (Francis et al., 2018). This PhD research explored the role of interoception in moral decision-making and moral behaviour across four studies implementing a range of physiological measures, including an immersive VR simulation of a moral dilemma.

The following literature review begins with a focus on interoception; including its neural bases (Craig, 2015), theoretical models of interoceptive processing (Ainley et al., 2016; Seth, 2013), measurement of interoceptive constructs (Garfinkel & Critchley, 2013) and overt behavioural responses associated with interoception. Empirical evidence and

theoretical accounts supporting a link between physiological and emotional arousal and moral decision-making are then discussed, followed by a review of more recent studies using virtual reality paradigms to explore harm-based moral behaviour. Following this, the research studies are presented within broader reflective commentary. Study 1 explored whether hunger, thirst, interoceptive sensibility and momentary emotional state influenced moral judgments of harm in an online cross-sectional design. Study 2 measured several parameters of physiological arousal and cardiac and gastric measures of interoception in an experimental lab study, to explore the link between arousal, interoception and moral judgments of harm. Study 3 investigated the relationship between physiological arousal, moral behaviour (harmful action) and interoception in immersive VR moral dilemma scenarios in a semi-autonomous vehicle. Study 4 investigated the effects of harm-salience and frame-type of coronavirus media articles and interoceptive sensibility, on behavioural intentions during the pandemic and moral judgments about the treatment of coronavirus patients.

Interoception

Interoception is fundamentally linked with the homeostatic regulation of bodily functions and refers to our perceptual experiences of internal events produced by the states and processes of our physiological body and organs (Craig, 2015; Farb et al., 2015). Interoceptive signals (inside the body) are different from exteroceptive signals (outside the body) and proprioceptive signals (signals relating to body position and movement) (Craig, 2015; Schulz, 2015). The integration of visceral signals with proprioceptive and exteroceptive signals gives rise to dynamic internal representations of the self and the external world (Farb et al., 2015; Seth, Suzuki & Critchley, 2012). Furthermore, increasing evidence suggests interoceptive signals are centrally important in generating subjective experiences of emotion and motivation (Craig, 2015; Feldman Barrett, 2017; Seth, 2013). Subjective feeling states are believed to result from interoceptive representations of affect-laden motivations; specifically, activity in the autonomic nervous system is proposed to elicit behavioural responses we concurrently 'feel' or experience (Craig, 2015). Interoception has shown to be directly and indirectly related to a range of psychological processes including attention and perception, cognition, memory and overt behaviour (for a review, see Farb et al., 2015; Tsakiris & Critchley, 2016). Although there is some evidence for a more general interoceptive ability that is consistent across visceral channels, such as relationships between heartbeat detection accuracy, gastric sensitivity (Herbert et al., 2012) and pain sensitivity (Pollatos et al., 2012), interoception is generally regarded as a multidimensional phenomenon

consisting of sub-constructs that may sometimes be interrelated but are largely independent (Ferentzi, et al., 2018a, 2018b; Garfinkel et al., 2015; Werner, Duschek, et al., 2009). There are a range of sensory receptors in the body which give rise to sensory information (Craig, 2015; Schulz, 2015):

1) Exteroceptors: e.g. mechanoreceptors in the skin, which process external sensory information. Occasionally body-internal events like heartbeats could be partially processed via exteroceptive somatosensory receptors.
2) Proprioceptors: e.g. the spindles of skeletal muscles. These receptors provide information related to body position, tension and muscle movement.
3) Interoceptors: these include a range of mechano-, chemo-, thermo- and metaboreceptors within the visceral organs which altogether constitute 'interoception' (Craig, 2015; Schulz, 2015).

Information from interoceptors is communicated to the central nervous system, with a large proportion transmitted via the brainstem or spinal cord below conscious perception (Craig, 2015). Interoceptive information is likely to reach consciousness if interoceptive information surpasses a fixed range of 'normal functioning' (Schulz, 2015). The nucleus tractus solutarius is the primary sensory brainstem for visceral-afferent information, with reflex circuits regulating homeostasis in physiological processes (Schulz, 2015). Brain areas implicated in the processing of interoceptive signals are the anterior cingulate cortex, anterior insula, somatosensory cortex and the prefrontal cortex, with the thalamus also responsible for the processing and perception of bodily signals (Schulz, 2015). The processing of interoceptive signals, like breathing or heartbeats can generate unique electrocortical potentials, such as heartbeat-evoked potentials (HEP's), which can be observed in the aforementioned brain regions (Pollatos et al., 2005; Schultz, 2015). Interoceptor subtypes can be distinguished by their functionality, for example, receptors may be responsive to pressure or movement (mechanoreceptors) or to temperature change above or below certain thresholds (thermoreceptors) (Schulz, 2015). They can also be characterised by their adaptability (i.e. how quickly they adapt to stimulation), and their sensory threshold, for example different subgroups of baroreceptors in arterial blood vessels are sensitive to changes in blood pressure within different ranges (see Schulz, 2015 for a review).

Predictive coding models of interoception

From a functional neuroanatomy perspective, Craig (2015) proposes the principle of optimal energy utilization is responsible for the evolution of humans' perceptual capacity for subjective affect and emotions. Evidence strongly suggests feelings we experience in the body represent discrete sensations integrally coloured by strong positive or negative affect (Feldman Barrett, 2017). These affect-laden sensations are fundamentally linked to motivations driving behavioural responses necessary to sustain the health and functioning of the body (Craig, 2015; Seth & Friston, 2016). In this way, the brain uses interoceptive information like a form of currency for homeostatic valuation i.e. weighing up the costs and benefits of any current or potential behaviours for the overall functioning of the body. This valuation process enables the most energy-efficient behavioural response (Craig, 2015).

The embodied nature of emotions proposed by Craig (2015) is a unifying premise of earlier neurophysiological theories of emotion. The James-Lange theory was the first to conceptualise emotional experience as fundamentally linked to the perception of changes in bodily state which are cognitively contextualised (Lang, 1994). Furthermore, Higgin's self-discrepancy theory (1987) hypothesised an adaptive basis of comparing sensed and expected sensations as a necessary process to bring our awareness toward unexpected changes in physiology which need to be attended to, as well as comparing current and predicted future states as a means to facilitate action. More recently, the Somatic Marker Hypothesis (SMH; Damasio, 1996) developed from research into emotion-based learning, suggests a central role for the ventromedial prefrontal cortex in processing somatic markers in the body. The SMH explains how afferent signals from visceral organs that are indicative of outcomes associated with reward and punishment, have the potential to bias cognitive processes, judgments and behaviours (Damasio, 1996). More recent interpretations of the link between physiological arousal and emotion emphasise that patterns of physiology in the body do not automatically map onto feeling states, rather, it is the sensitivity to, or perception of, subtle somatosensory signals that contribute to subjective experiences of affect (Feldman Barrett et al., 2004; Feldman Barrett, 2017). For example, individuals more attuned to their internal processes report more intense emotional responses (Wiens et al., 2000).

Functional accounts of the predictive power of interoceptive signals to influence decision-making and motivate behaviour can be assimilated with computational neuroscience models of the 'Bayesian Brain' (Friston, 2010). Bayesian perspectives describe neural

processes as inherently statistical, as the brain produces probability-driven hypotheses for stimuli encountered in the environment, which are then compared with incoming sensory information (Friston, 2010; Seth, 2013; Seth & Friston, 2016). This theory has been applied to interoceptive signalling, hypothesising how Bayesian probability distributions about states in the world give rise to internal sensations. 'Hypotheses' and 'beliefs' are proposed to represent neuronally encoded probabilities about the internal or external causes of sensory information. More precise probability distributions are likely to have more inferential power. Causes are 'hidden' as they can only be inferred from sensory consequences of hypothesised states in the world, and therefore can never be objectively perceived (Ainley et al., 2016; Feldman Barrett, 2017; Friston, 2010; Seth & Friston, 2016; Seth, 2013).

The predictive coding model (PCM) of interoception (Seth, 2013) describes how 'simulation maps' in memory represent the integration of sensory signals across the body which contribute to a holistic representation of the self that can interact with cognitive processes (Farb et al., 2015; Seth & Critchley, 2013). Predictive simulation maps provide a 'best guess' estimation of the state of the body at a given moment which is based on prior states in memory (Seth, 2013; Farb et al., 2015). Therefore, these maps represent an interpreted signal which is achieved through the integration of the range of afferent sensory signals occurring at one point in time (Seth, 2013; Seth & Friston, 2016). Subjective emotion represents a quality of sensory representations that are accompanied by interoceptive predictions and interpreted alongside exteroceptive and proprioceptive signals (Feldman Barrett, 2017; Seth, 2013).

Within these simulation maps, motivational salience is represented as the emotional valence we experience (Farb et al., 2015). Negatively valanced signals signify a deviation away from a pre-defined homeostatic threshold e.g. during hunger, which must be remedied through physiological or behavioural change - the process of 'allostasis' (Farb et al., 2015; Sterling, 2015). Allostasis involves a range of functions including autonomic, neuroendocrine and behavioural processes, with many regulatory physiological processes happening automatically and outside conscious awareness (Craig, 2015; Farb et al., 2015; Sterling, 2015). For example, a cold glass in freezing temperatures will feel unpleasant, whereas it will be received gratefully in the summer when our bodies are trying to cool down. These feelings of pleasantness or unpleasantness represent the concurrent perception of our behavioural motivation to continue holding the cold glass or let it go (Craig, 2015; Critchley & Harrison, 2013).

At lower levels, the simulation map consists of 'raw' sensory afferents with higher levels aggregating information from the lower levels of the map into representations that could be consciously available (Farb et al., 2015). What is consciously available in the simulation map is proposed to differ between people, depending on factors such as genetics, goals or cultural experiences (see Farb et al., 2015; Maister & Tsakiris, 2014; Ma-Kellums, 2014). Regulatory motivation is believed to be influenced by the extent to which unexpected sensations that deviate from an expected body state (the prior) are perceived as acceptable or problematic (Farb et al., 2015; Seth & Friston, 2016). To resolve prediction error that results from the disagreement between expected and experienced sensations, the simulation map may update itself at that layer of the map (perceptual inference) or adjust the internal sensations (active inference) (Ainley et al., 2016; Friston, 2010; Seth & Friston, 2016). Larger prediction errors are more likely to reach conscious awareness and engage 'higher-order' cognitive processing and overt behavioural responses (Craig, 2015; Critchley & Garfinkel, 2018; Farb et al, 2015). Farb et al (2015) suggest conscious deliberation may result in the evaluation of several potential regulatory strategies to minimize prediction error.

Active versus perceptual inference

According to PCM models of interoception (Seth, 2013), prediction error can lead to an active or perceptual form of inference (Ainley et al., 2016; Farb et al., 2015). In the case of active inference, resolving prediction error is achieved by altering our interoceptive state(s) to be congruent with the expected state (prior), thus initiating a regulatory response aimed at altering interoceptive processes to be more in line with the expected state. Seth and Friston (2016) propose sensory faculties may then be activated to re-examine sensory information and selectively sample this information as evidence to support the expected state. Active inference often happens outside our conscious awareness, triggering physiological or hormonal changes, not requiring overt behavioural regulation (Farb et al., 2015; Gu & FitzGerald, 2014). In contrast, perceptual inference involves altering the expected state (prior) to correspond with current interoceptive sensations (Seth & Friston, 2016; Ainley et al., 2016), thus, prioritising the accuracy of sensory information (Farb et al., 2015). In this instance, our simulation maps are updated to more precisely represent our current internal sensations (Seth, 2013). Farb et al (2015) proposes that perceptual inference may support the widening of sensory expectations which could reduce reactivity towards surprising interoceptive sensations. They further suggest that overt active inference could lead to dysregulation of interoceptive signals in the interests of a regulatory response aimed at

resolving prediction error. For example, for highly stressed individuals, prior expectations may be favoured over perceptual inference to readily identify negative sensations and initiate behaviours that reduce those sensations quickly (Farb et al., 2015).

In contrast, Farb et al (2015) propose that perceptual inference may represent a more adaptive response to surprising interoceptive information, if interoceptive representations are allowed to occur and diminish, without necessarily committing to rigid appraisals of the sensations experienced; a process described as 'decentering'. Contemplative training and traditions such as mindfulness promote practices of decentering which encourage acceptance of interoceptive experiences as a way to regulate interpretations of surprising interoceptive sensations (Mikulas, 2011), rather than engaging more active emotional regulation processes to manage these sensations (Gross, 2002; Farb et al., 2015). Interestingly, increased cortical thickness of brain regions associated with interoceptive processing has been found in long-term meditators with extensive insight meditation experience compared with matched controls (Lazar et al., 2005). However, there is little evidence to suggest the 'benefits' of contemplative practice can lead to enhanced interoceptive accuracy (Ferentzi et al., 2018b). No difference in interoceptive accuracy in a heartbeat detection task was found between Buddhist meditators and non-meditators (Khalsa et al., 2008), and many contemplative practice interventions have been unsuccessful at improving interoceptive abilities such as heartbeat detection accuracy (see Ferentzi et al., 2018b for a review).

Measures of interoception

Garfinkel & Critchley's (2013) hierarchical model of interoception is currently regarded as the most plausible and well-conceptualised account of the distinct dimensions of interoception. The model distinguishes between: *Interoceptive accuracy/sensitivity*: objective performance on tests of interoceptive monitoring e.g. heartbeat detection and counting (Schandry, 1981); *Interoceptive sensibility*: A dispositional tendency to focus internally which is assessed via self-report questionnaires and; *Interoceptive awareness*: a meta-cognitive awareness of interoceptive accuracy, that is, whether our confidence in our accuracy reflects actual performance (Garfinkel & Critchley, 2013). Research has consistently demonstrated the independence of these dimensions (see Garfinkel & Critchley, 2013; Ferentzi et al., 2018b; Whitehead et al., 1977). However, relationships between these facets of interoception and their respective influences on behaviour are less clear, as studies have often limited their measures of interoception to one or two dimensions.

Interoceptive sensibility and awareness

Self-report measures of interoceptive sensibility and awareness are independent from measures of interoceptive accuracy which refer to individuals' objective performance on interoceptive tasks (Garfinkel et al., 2015; Garfinkel & Critchley, 2013). Interoceptive sensibility refers to a self-evaluated trait which can be measured by questionnaires, assessing an individuals' level of engagement with interoceptive sensations or belief in their interoceptive ability (Garfinkel et al., 2015). Questionnaires measuring interoceptive sensibility include the Autonomic Perception Questionnaire (Mandler et al., 1958), Body Perception Questionnaire (Porges, 1993) and the Body Consciousness Scale (Miller, Murphy, & Buss, 1981). These measures are useful to understand individual differences in subjective experiences of interoceptive sensations and attentional styles associated with bodily sensations. 'Body awareness' is a term often used to refer to self-reported sensitivity to bodily states which may be considered equivalent to interoceptive sensibility. Interoceptive sensibility does demonstrate good test-retest reliability but appears somewhat susceptible to external factors influencing bodily activity (e.g. Rani & Rao, 1994; see Ferentzi et al., 2018b). For example, contemplative training (Bornemann et al., 2015), mindfulness training (Fissler et al., 2016) and mind-body experience (Mehling et al., 2013) have been shown to influence interoceptive sensibility. Ferentzi et al (2018b) suggest that mind-body interventions emphasising the importance of listening to bodily sensations may promote self-reported awareness of such sensations.

Subjective accounts of interoceptive ability are likely to provide a biased estimation of interoceptive ability and are not necessarily predictive of objective measures of interoception (Garfinkel et al., 2015). A measure combining both interoceptive sensibility and interoceptive accuracy involves a subjective confidence score for performance on interoceptive tasks, such as heartbeat detection tasks, which can be compared with performance accuracy on that task (Ehlers et al., 1995; Garfinkel et al., 2015; Khalsa et al., 2008). This measure represents individuals' metacognitive awareness of interoception and is the most well-conceptualised definition of interoceptive awareness (Garfinkel et al., 2015; Garfinkel & Critchley, 2013). A more recently validated measure- The Interoceptive Accuracy Scale- developed by Murphy and colleagues (2018), captures self-perceived accuracy of interoception across a range of visceral systems not limited to accuracy on heartbeat detection tasks. Despite these important differences very few studies have

distinguished between heartbeat detection accuracy and individuals' awareness or confidence in their interoceptive ability.

Interoceptive accuracy

Cardiac system

Cardiac interoception can be measured using heartbeat perception tasks to understand the link between heartbeats and cognition and have generally been the principal method for evaluating interoceptive accuracy (Critchley & Garfinkel, 2018). Baroreceptors in arterial blood vessels are responsible for transmitting information concerning cardiovascular arousal to the brain (Craig, 2015; Critchley & Garfinkel, 2018). Heartbeat counting tasks (HBC) and heartbeat detection or discrimination (HBD) tasks represent the two methods for assessing cardiac interoceptive accuracy (Garfinkel et al., 2015; Ring & Brener, 2018). For heartbeat counting tasks participants are required to count their heartbeats for indicated time intervals and accuracy is determined by calculating the difference between reported and actual heartbeats (Schandry, 1981). Heartbeat counting accuracy can be calculated using two slightly different formulae: Schandry's (1981) calculation (scores range between 0-1) and Hart et al's (2013) calculation (scores range between -1 and 1) which accounts for overestimations as well as underestimations of heart beats. Some researchers have compared heartbeat perception across conditions of sitting, standing and supine positions, and post-exercise (see Ring & Brener, 2018 for a review) while others instruct participants to maintain a seated position (e.g. Herbert et al., 2012).

The validity of the heartbeat counting method has been challenged by evidence showing heartbeat counting is more dependent on beliefs about (resting) heart rate than heartbeat sensations themselves (Ring et al., 2015; Ring & Brener, 2018). In addition, research has shown individuals are capable of estimating their heart rate without experiencing their heart rate and are therefore able to perform accurately on the task (e.g. Kleckner et al., 2015). However, predictive coding model perspectives of interoception (Ainley et al., 2016; Farb et al., 2015; Seth & Friston, 2016) suggest heart rate beliefs are likely to be informed by a lifetime experience of heart rate sensations and thus prior expectations about heart rate are likely to be generated as a result of both explicit and tacit knowledge about heart rates, which may be associated with performance on heartbeat counting tasks.

Heartbeat discrimination tasks (HBD) are believed to be the most valid measure of cardioceptive accuracy (Brener et al., 1993; Schneider, Ring, & Katkin, 1998). Performance on these tasks is not influenced by prior beliefs or knowledge about heart rate and appears to rely completely on perception of heartbeat sensations in the moment (Brener & Ring, 2016; Ring & Brener, 2018). HBD tasks require participants to make judgments about whether their heartbeats coincide with exteroceptive stimuli (e.g. acoustic tone) which are presented at time points that coincide and do not coincide with participants' real heartbeats (Whitehead et al., 1977). Participants are attached to an electrocardiogram (ECG), which monitors cardiac activity. Stimuli delays are presented following the R wave of the ECG cycle (Whitehead et al, 1977). Evidence suggests that around 200ms and 500ms represent the optimum R-wave delay for distinguishing between coincident (200ms) and non-coincident tones (500ms) (see Kleckner et al, 2015). Precision is recorded across a large number of trials which can be interpreted as heartbeat detection accuracy (Ring & Brener, 2018). Kleckner et al (2015) recently found that fewer than forty trials will not yield sufficient power or reliability on this task, suggesting potentially high experimental load for participants.

There is some evidence showing correlations between performance on the HBC and HBD tasks (e.g. Hart et al., 2013; Knoll & Hodapp, 1992). However, a recent comparison found no relationship between these tasks (Ring & Brener, 2018). Ring and Brener (2018) concluded these tasks measure different abilities and advise against using the HBC task - despite performance on heartbeat counting tasks predicting a range of psychological phenomena (Ainley et al., 2016; Tsakiris & Critchley, 2016). In contrast, others have suggested that the HBD task may not be suitable for understanding individual differences in interoceptive sensitivity because of the difficulty of the task (Kleckner et al., 2015). Furthermore, Forkmann et al (2016) proposed that the HBD and HBC tasks require different psychological processes. Whereas the tracking task requires attention to be focused completely on visceral sensations, the HBD task asks people to pay attention to both external and visceral sensations at the same time, which (unlike the HBC task) demands effective multi-sensory integration of this information to perform well.

There is considerable evidence to suggest that interoceptive accuracy and interoceptive sensibility are not related (e.g. Ainley & Tsakiris, 2013; Critchley et al., 2004; et al., 2017; see Ferentzi et al., 2018b for a review). A recent longitudinal non-intervention study also found no predictive relationship between body awareness and interoceptive

accuracy (Ferentzi et al., 2018b). However, Duschek et al (2015) found participants with higher cardiac interoceptive accuracy reported greater interoceptive sensibility using the Body Consciousness Questionnaire (Miller, Murphy, & Buss, 1981). The authors argue that failure to find an association between cardiac interoceptive accuracy and private body consciousness could be explained by the relatively 'low' interoceptive ability of previous samples based on proposed interoceptive criterion (Schandry, 1981). However, many studies use slightly different methods of heartbeat counting and discrimination tasks, and those that operationalise heartbeat counting accuracy as a dichotomous variable may use different cut-off points to distinguish between 'high' and 'low' interoceptive accuracy, depending on the level of interoceptive accuracy within the sample.

Gastrointestinal system

Historically, interoception research has largely focused investigations on the cardiac system which has resulted in a limited understanding of the relationships between interoceptive constructs across visceral systems. Another interoceptive system which can be measured using standardized interoceptive tasks is the gastrointestinal (GI) system (Herbert et al., 2012). Previous work investigating interoceptive GI processes have traditionally used invasive measurement techniques (e.g. manipulating distention of stomach) which can affect GI processes themselves. A more invasive method for measuring perceptions of stomach contractions involves a perfused catheter that delivers water into participants' stomach via a tube through their nose. Stomach contractions are measured, and participants indicate whether a stomach contraction co-occurs with a light (Whitehead & Drescher, 1980). Invasive tasks like this have been criticised for increasing the likelihood of learning effects, whereby participants awareness of sensations associated with gastric functioning are enhanced during the task. Moreover, tolerance of this unusual procedure is likely to be restricted to certain individuals which is likely to compromise the representativeness of sample (Herbert et al., 2012).

A less invasive method for assessing perception of gastric sensations is a standardized water load test (WLT) which activates gastric distension and post-ingestion gastric neuromuscular activity (see Herbert et al., 2012). Water loads have shown to stimulate normal slow-wave frequency of the stomach (e.g. Koch et al., 2000), and using water avoids the hormonal consequences associated with eating food (see Herbert et al., 2012). The WLT has shown to be a valid measure of subjectively felt fullness and has been consistently

replicated in healthy participants and patients with gastrointestinal disorders (e.g. Jones et al., 2003; Koch et al., 2000; Van Dyck et al., 2016; Herbert et al., 2012). Water load tasks also demonstrate good correlation with previous invasive methods (e.g. Boeckxstaens et al., 1990).

Two studies that have explored perception of visceral sensations from both the cardiac (using heartbeat counting method) and GI systems, have found correlations between interoceptive sensitivity of these systems (e.g. Whitehead & Drescher, 1980; Herbert et al., 2012). However, a recent study did not find correlations across different sensory modalities of interoception, including gastric perception, pain, balance, proprioception, and balance and concluded that interoceptive sensitivity cannot be generalised across visceral systems (Ferentzi et al., 2018a). Importantly, they used a water drinking task which was fundamentally different to the prior studies (Herbert et al., 2012; Whitehead & Drescher, 1980), whereby participants drank a height-corrected amount of water per minute for five minutes and reported sensations of fullness and unpleasantness. In contrast, the WLT used by Herbert et al (2012) involved drinking water freely for 5 minutes until reaching a point of maximum fullness, along with reporting other sensations such as nausea and satiety. More recently, Van Dyck et al (2016) developed a two-stage water load task, whereby participants drink water until the point of satiation (1st threshold) and to the point of maximum fullness (2nd threshold). This allows the calculation of a percentage amount of water required to achieve maximum fullness from the point of satiety, resulting in a measure of gastric interoception that is not determined by stomach capacity (Van Dyck et al., 2016).

Interoceptive processing

Neuroimaging techniques allow the comparison of interoceptive predictions with subjective experiences during the recording of brain activity. The anterior insula (AI) cortex is implicated in interoceptive processing and has shown to be activated when people are paying attention to their internal bodily state (Terasawa et al., 2012). The AI is also believed to be central in facilitating predictive representations of emotion and uncertainty, and also empathy, which can influence decision-making (Singer et al., 2009). The middle insula appears to be involved in the integration of interoceptive information with exteroceptive context (Seth & Friston, 2016). Interoceptive processing can also be observed as event-related potentials, such as heartbeat-evoked potentials (HEP's) using electroencephalograms (EEG; Schulz, 2015). HEP's correspond with the perception of cardiac signals but do not

necessarily represent conscious perception of heartbeats (Schulz, 2015). HEP's can be measured alongside heartbeat detection tasks (e.g. Leopold & Schandry, 2001) and have shown to predict heartbeat detection performance, attentional focus on heartbeats and motivation to perform in heartbeat detection tasks (see Schulz, 2015 for a review). Respiratory-related evoked potentials (RREP's) are a less well-studied measure, but respiratory occlusion methods have shown to induce respiratory-related evoked potentials captured using EEG (e.g. von Leupoldt et al., 2010). Finally, the effects of interoceptive processing on the preconscious processing of external information can be measured as the cardiac modulation of startle response (Schulz et al., 2016), as startle responses are lower during the earlier phases of the cardiac cycle (Schulz, 2015). This is the only method that is able to capture the influence of afferent interoceptive signals communicated via the brainstem (Schulz, 2015).

Temporal stability of interoception

There is some evidence that interoceptive abilities may be susceptible to change. For example, contemplative practices such as mindfulness can improve self-reported interoceptive sensibility (Bornemann et al., 2015) and age has been associated with reduced heartbeat detection ability (Khalsa et al., 2009). However, other evidence suggests interoceptive-accuracy measured by heartbeat detection tasks is a relatively stable construct. Parkin et al (2014) found heartbeat counting accuracy remained relatively consistent after a 1-week and 8-week mindfulness intervention. Furthermore, in a longitudinal non-intervention study Ferentzi et al (2018b) found interoceptive accuracy (on a heartbeat counting task) demonstrated good test-retest reliability over time. Providing feedback to participants on their performance in heartbeat detection and counting tasks has been used to improve interoceptive accuracy (e.g. Meyerholz et al., 2019). However, Ferentzi et al (2018b) suggest it is unclear whether performance enhancement is the result of updating heart rate beliefs or genuine improvement in accuracy (Ring et al., 2015). Other intervention studies using chemical manipulations (isoproterenol) to increase the heart rate found changes in participants reported cardiac sensations according to the dose (Khalsa et al., 2009). The impact of stress has also been shown to influence heartbeat detection for females, with accuracy declining during a challenging cognitive task (Fairclough & Goodwin, 2007).

Other evidence suggests interoceptive accuracy may be improved when combined with exteroceptive cues. For example, Ainley et al (2013) found heartbeat detection ability

was significantly enhanced when participants were able to look at their own face in a mirror. In addition, Suzuki et al (2013) used a 'cardiac rubber hand illusion' which integrated interoceptive information about heart rate into a computer-generated augmented reality. They found virtual hand ownership was increased when cardio-visual feedback was synchronous with the participants' heartbeat and this correlated with interoceptive sensitivity. These studies highlight the importance of congruence between interoceptive and exteroceptive information in the representation of internal states, which may influence momentary interoceptive abilities.

Interoception and behaviour

In line with the predictive coding model of interoception (Seth, 2013; Seth & Friston, 2016), comparing incoming sensory information with prior expectations results in prediction error which motivates modifications to either the bodily signals or the simulation map through autonomic and behavioural regulation. Although these self-regulatory mechanisms are often not subject to conscious awareness, interoceptive signals can be powerful enough to produce overt behavioural responses (Craig, 2015; Farb et al., 2015). There is evidence to support the notion that our behaviour may be driven by our allostatic drives; for example, holding a hot cup of coffee has shown to increase feelings of interpersonal warmth beyond a change in mood (Williams & Bargh, 2008), and holding a cold object has shown to decrease interpersonal trust (Kang et al., 2011). Importantly, not all behaviour is an attempt to achieve balance in our bodies and Farb et al (2015) suggest that rather than regulatory motivation being driven by allostasis, it may be more usefully conceptualised as how our sensations correspond with predicted or desired states.

Enhanced interoceptive accuracy in particular has shown to facilitate more adaptive self-regulation and behavioural responses (e.g. Ainley et al., 2016; Dunn et al., 2010; Füstös et al., 2013). Ullsperger et al (2010) propose interoceptive awareness is essential for conscious error perception and subsequent behavioural adjustments due to the involvement of the anterior insular cortex. Interestingly, attentional resources are influenced by cardiac signals, for example afferent signals sent during the heart contraction phase (systole) decrease distraction towards background stimuli and enhance performance on visual search tasks (Pramme et al., 2016). In addition, systole has been shown to selectively enhance the detection of fear faces in an emotional attentional blink paradigm (Garfinkel & Critchley, 2016). Furthermore, those with better interoceptive accuracy on a heartbeat detection task

show better implicit memory (Werner et al., 2010) and decision-making ability on intuitive reasoning tasks (Dunn et al., 2010). Herbert et al (2007) also found better heartbeat perceivers showed enhanced self-regulation of physical load and cardiovascular effort during exercise. Moreover, Weiss et al (2014) found greater interoceptive accuracy during a heartbeat counting task was associated with more effective self-regulation and pain tolerance - although a relationship between pain tolerance and interoceptive accuracy was not found by Werner and colleagues (2009). Using EEG, Füstös et al (2013) found interoceptive accuracy supported downregulation of emotional arousal during an emotional reappraisal procedure. Relatedly, Kever et al (2015) found heartbeat detection ability was associated with both reappraisal and suppression focused emotional regulation strategies. These findings suggest that a superior ability to detect cardiac sensations, may support more effective and dynamic use of emotional regulation strategies in the moment (Kever et al., 2015), which could have benefits for wellbeing and behavioural responses to emotional stimuli.

There is some evidence to suggest that interoceptive accuracy may not always support optimal behavioural responses. For example, Dunn et al (2010) found that changes in participants' heart rate distinguished between 'profitable' and 'unprofitable' decks during a computerised gambling task. When these somatic markers were indicative of advantageous choices, higher interoceptive accuracy corresponded with more successful intuitive decision-making. However, interoceptive accuracy was problematic for intuitive learning of profitable versus unprofitable decks when bodily signals were more suggestive of disadvantageous decisions. More recently, Marshall et al (2018) proposed the integration of interoceptive signals into motor representations could lead to more impulsive behaviour in certain contexts. In support of this, Ainley et al (2014) found that good heartbeat perceivers found it more difficult to inhibit the tendency to imitate the actions of others, which they suggested may be associated with enhanced empathy among these people. Marshall et al (2018) suggested greater interoceptive accuracy could engender stronger internal representations of actions, which leads to increased motor reactivity to observed actions.

At a more fundamental level, interoceptive signals appear to be important in shaping our perception of time (Di Lernia et al., 2018), and there is increasing evidence to suggest an association between the physiological regulation of the body and conscious awareness of our embodied self (e.g. Moseley et al., 2008; Seth, Suzuki & Critchley, 2012; Seth & Friston, 2016; Farb et al., 2015). For example, brushing an embodied rubber hand (using the Rubber

Hand Illusion paradigm) has been shown to decrease the temperature of participants' real hands (Moseley et al., 2008). Interoception is also proposed to be fundamentally linked with experiences of agency (Nahab et al., 2011; Seth, Suzuki & Critchley, 2012) and interoceptive processes have been implicated in experiences of empathy (e.g. Fukushima, Terasawa, & UmedaIf, 2011). These findings highlight how low-level visceral processes may interact with higher-level social processes such as our sense of self, which could influence more complex social behaviour. Importantly, there are cultural variations in conceptualisations of interoception, interoceptive awareness and cultural practices that encourage an awareness of somatic states and bodily sensations (Ma-Kellams, 2014). So far, these differences have typically been examined in comparisons of Eastern and Western cultures, but further cross-cultural research is needed to illuminate the nuances in how we understand interoception such as the language we use, differences in interoceptive ability and cultural activities and traditions that shape our relationship to interoceptive sensations (Ma-Kellams, 2014; Tsakiris, 2020).

Somatic markers, moral judgments and behaviour

Moral norms can vary greatly across cultures, religions and political groups but norms surrounding principles of harm and fairness appear to be recognised more consistently than others such as those relating to purity or authority (Haidt & Graham, 2007). Our sensitivity to harm including observations of aggression and cruelty, originated from an evolutionary response to the suffering in others, and is upheld in moral virtues of compassion and kindness (Haidt & Graham, 2007). Historically, moral psychology and moral philosophy focused more heavily on the cognitive and contextual factors influencing our evaluation of 'right' and 'wrong' in hypothetical text-based moral dilemmas (see Teper et al., 2015 for review). Over time, there has been increasing impetus to understand how emotional and physiological factors may interact with cognitive processes to influence moral judgments and behaviour particularly in relation to harm-based moral norms (Blair, 1995; Buon et al., 2016; Cushman et al., 2012; Damasio et al., 1990; Greene et al., 2001, 2004, 2009; Haidt, 2001; Parton & McGinley, 2019; Reynolds & Conway, 2018).

In a situation of moral conflict, a negative emotional response could signal the anticipation of taking a moral action that is inconsistent with our principles (e.g. Cushman et al; 2012; Greene et al., 2001). Feelings of frustration, pain and anxiety are often reported by people who have to make difficult decisions that are incongruent with their ethical principles

(Corley, 2002). Evolutionary perspectives suggest that the emotional aversion we experience when carrying out or anticipate carrying out moral transgressions formed due to our need for human cooperation to survive (Rand & Nowak, 2013; Teper et al., 2015). Negative affective experiences have been proposed to represent outcome-based aversion mechanisms which are adaptive because they facilitate moral decisions that have beneficial consequences for ourselves and societal functioning in the long term, such as mutual cooperation and respect (Rand & Nowak, 2013). However, when making moral decisions in complex social environments, we may not always have the time nor all the necessary information to weigh up the consequences for our future selves in a rational and deliberative manner (Damasio, 1996; Kahneman, 2003). The Somatic Marker Hypothesis (SMH; Damasio, 1996) proposes that decisions, including moral decisions, are influenced by automatic 'intuitive' inclinations about the 'rightness' or 'wrongness' of a given decision which is informed by somatic signals generated in the body.

The SMH has received considerable empirical support in moral judgment studies (Koenigs et al., 2007; Moretto et al., 2010; Young et al., 2010). Damage to brain areas involved in the processing of physiological experiences have been associated with anti-social behaviour over time (Damasio et al., 1990). Measures of anti-social behaviour have been shown to predict reduced heartbeat detection accuracy (Nentjes et al., 2013). The authors suggest that a reduced sensitivity to interoceptive sensations could influence a likelihood to carry out anti-social behaviour, but they acknowledge cause and effect are difficult to establish (Nentjes et al., 2013). It is possible that poorer interoceptive accuracy could indicate a reduced ability to self-regulate which may predispose people towards acting anti-socially, rather than interoceptive accuracy representing an inability to weigh up the moral implications of behaving anti-socially. Furthermore, there is evidence to suggest that the simulation of harmful actions is experienced as emotionally aversive before we contemplate the consequences of that action (e.g. Blair, 1995; Cushman et al., 2012) and these negative emotions are what prevent us from carrying out transgressions (Miller et al., 2014). Cushman et al (2012) found that individual differences in anticipatory physiological arousal influenced readiness to carry out simulated harmful actions (e.g. hitting a plastic baby doll) compared to when participants observed someone else carry out those actions. This suggests we physiologically respond to the anticipation of carrying out a harmful behaviour despite knowing that there are no harmful consequences of the action. Differences in interoceptive capacities can be conceptualised as individual difference measures of our ability to perceive,

or notice somatic signals (Garfinkel et al., 2013) but it is currently unclear how individual differences in interoception could influence the relationship between emotional arousal and harm-based moral decision-making.

There is some evidence to suggest that a belief that we are physiologically aroused can influence moral decision-making. In a series of laboratory experiments, Gu et al (2013) explored the influence of digitally altered cardiac feedback on moral behaviour. They provided false heartbeat feedback ('fast' versus 'normal') to participants while they contemplated a moral action: either volunteering time for a charity or deceiving a confederate. They found 'fast' versus normal heartbeat feedback increased volunteering rates and reduced deception. Differences in moral behaviour was accounted for by stress and coping theory (Lazarus & Folkman, 1984) which proposes perceived stress (i.e. negative physiological arousal) during moral conflict activates coping mechanisms aimed at regulating negative emotions and/or removing the stressor. This could result in either conforming to moral norms (in this case volunteering for charity) or implementing moral disengagement strategies (Bandura, 2002; Gu et al., 2013). In addition, perceived heartbeat impacted moral action less when individuals were mindful and when approaching the situation deliberatively. Gu et al (2013) concluded that trait mindfulness and deliberative processing buffered participants' responses to cardiovascular stress signals. Conversely, Gu et al (2013) speculated that individuals low in interoceptive sensitivity may be less influenced by a perceived 'fast' heartbeat as they have less knowledge of their cardiac sensations in stressful situations. However, as interoception and physiological arousal was not measured, it is unknown whether the effects they found were the result of real physiological changes.

Moral dilemmas

Moral dilemma thought experiments are useful for studying our aversion to moral transgressions because the type and means of carrying out immoral acts can be systematically varied. A moral dilemma typically offers the choice between two 'wrong' courses of action, in that both choices involve the violation of one or more moral principles - hence the dilemma. Two fundamental philosophies proposing conflicting maxims for determining whether actions are morally right are John Stuart Mill's Utilitarianism (Mill, 1998) and Immanuel Kant's duty-based (deontological) ethics (Kant, 2018). Mill's Utilitarianism (Mill, 1998) proposes moral actions that maximise the wellbeing of the greatest number of people are good actions. Whereas Kant's deontological ethics assesses the moral 'goodness' of an

action based on the act itself and how it relates to moral rules pertaining to the rights and duties of those involved in the situation.

The well-known and widely used thought experiments in moral psychology that tap into our aversion to harm are the Trolley (Thomson, 1985) and Footbridge dilemmas (Foot, 1967); which put deontological and utilitarian principles in competition with one another. The Trolley dilemma requires participants to judge whether it is morally right to switch a driverless trolley travelling towards five people (who would certainly be killed) to an alternate track with one person on it - saving five people but killing one. Whereas the Footbridge dilemma asks participants whether it is morally right for them to push a large man standing near them off the footbridge to stop the trolley from killing five people on the track but killing the man in the process. These dilemmas distinguish between 'personal' and 'impersonal' harmful actions (Greene et al., 2001). The Footbridge dilemma requires either anticipated or actual physical force and death is necessary to achieve a utilitarian outcome. Whereas harm caused in the Trolley dilemma is done impersonally and death is a side-effect of switching the trolley. The outcome of each action is the same, but people tend to endorse a utilitarian response in the Trolley dilemma, but not in the Footbridge dilemma (e.g. Cushman et al., 2006; Greene et al., 2001, 2004, 2009; Moretto et al., 2010; Patil, 2015), showing how the mental representation of harmful acts and not just their consequences, can influence people's moral judgments (Mikhail, 2007).

Importantly, research has also uncovered an omission bias in moral dilemmas; whereby omissions of actions that lead to harmful consequences are judged as less morally wrong than moral actions that result in the equivalent amount of harm (e.g. Cushman et al., 2006; Navarrete et al., 2012). Relatedly, research suggests that a motivation be consistent in our moral responses (Lombrozo, 2009) can be overridden, when a less emotionally arousing situation is presented before more arousing situation as it has less influence on the more arousing situation (Schwitzgebel & Cushman, 2012). Essentially, if it is possible to revise our initial response with explicit reasoning, we may be less likely to be consistent in our moral judgments (Schwitzgebel & Cushman, 2012).

Physiological responses to harmful action

The physical proximity between the agent doing the harm, and the subsequent 'harm' carried out characterises the key difference between the Trolley and Footbridge dilemmas.

The terms ‘Personal’ and ‘Impersonal’ are also used to describe the difference between these dilemmas. However, Greene et al (2009) proposed ‘personal force’ offered a more precise experimental variable, which describes the extent an agent’s muscles and body are (hypothetically or not) involved in inflicting harm to another person (see Christensen & Gomila, 2012 for a review). Dual process models of moral behaviour describe how systematic and deliberative processes can interact with automatic and emotional processes when responding to hypothetical dilemmas of harmful moral action (e.g. Greene et al., 2001; Greene et al., 2004; Greene, 2007). The Trolley Dilemma is believed to allow a more cognitive and ‘cooler’ decision making process which leads to greater utilitarian judgments (Greene et al., 2001; Greene et al., 2004). Whereas the Footbridge dilemma requires us to imagine intentionally harming another person with our own body which facilitates a stronger aversive emotional response to the harmful action (Cushman, 2013; Cushman et al., 2006; Greene et al., 2001). This theory was primarily supported by a series of fMRI experiments, whereby Greene et al (2001) found personal moral dilemmas (e.g. the Footbridge dilemma) evoked a negative emotional response not present in impersonal moral dilemmas (e.g. the Trolley problem).

Perhaps unsurprisingly, physiological arousal is also greater during moral dilemmas where personal harm is required to attain a utilitarian outcome (McDonald et al., 2017; Moretto et al., 2010; Navarette et al., 2012). Greene et al’s (2001) dual process account suggests that condoning harm in personal dilemmas is only possible when the emotional processing system deterring us from causing harm is overridden by deliberative processing mechanisms which assess the costs and benefits of the action (Greene et al., 2009). In support, individual differences in arithmetic reflection, which potentially promotes the weighing up of ‘outcomes’ of harm in moral dilemmas (Baron et al., 2015) and cognitive reasoning (Patil et al., 2020) have been associated with utilitarian judgment. People are also more likely to reject harm when given less time to make a moral judgment, which the authors concluded was evidence that ‘cognitive control’ can be facilitated when more time for deliberation is allowed (Suter & Hertwig, 2011).

Conversely, Patil and Silani (2014) found trait alexithymia (a characteristic associated with an inability to identify emotions and deficits in empathy) predicted increased utilitarian responses to emotionally aversive personal moral dilemmas. The authors concluded that this finding reflected a reduced empathic response for the victim in alexithymic participants,

which supported enhanced cognitive processing and cost-benefit analysis. Combined, these findings suggest that an absence of an emotional response or an increase in cognitive deliberation, could both predict stronger utilitarian preferences. Greene et al (2004) found that compared with deontological judgments, utilitarian judgments are associated with greater activation in the dorsolateral prefrontal cortex - a region known to be important for cognitive control. Greene et al (2008) also found cognitive load selectively increased response time for utilitarian decisions but not for deontological decisions in personal moral dilemmas. However, Moore et al (2008) found that although people with greater working memory capacity were more likely to condone personal harm when harm was inevitable, their results suggested this was more due to enhanced deliberative reasoning as opposed to executive functioning 'overriding' or constraining emotional responses.

A reputational adaptation account of moral behaviour suggests automatic negative responses to inflicting harm with personal force may be an adaptive neurobiological response to deter us from actions where we are more likely to be judged as culpable (e.g. DeScioli et al., 2011). However, Greene (2009) suggested that there could be a variety of reasons why personal moral dilemmas generate an aversive emotional response, which may include: the personal or direct quality of the act (Greene et al., 2001; Royzman & Baron, 2002); the intentional nature of the action or the physical contact required (Cushman et al., 2006); the intervention required on the victim (Waldmann & Dieterich, 2007); or the culpability associated with personal harm (DeScioli et al., 2011). More recently, researchers have found stable individual differences in the emotional responses people have to carrying out harmful acts, and the consequences of harmful acts. The concept of 'Action aversion' describes individual differences in the negative emotional response to carrying out 'harmful' actions e.g. hitting someone with a rubber hammer (Cushman et al., 2012; Miller et al., 2014). 'Outcome aversion' represents another individual-differences measure associated with a negative reaction to witnessing harm, regardless of its cause (Miller et al., 2014). Action versus outcome aversion are proposed to represent distinct 'self' versus 'other' oriented motives respectively, which could influence moral judgments of harm (Miller et al., 2014).

Action-based harm aversion has been linked to physiological threat responses (Cushman et al., 2012), evidenced by increases in systemic vascular resistance (SVR; Mendes et al., 2007) and pre-ejection period (Parton & McGinley, 2019). Individuals with greater cardiovascular 'threat reactivity' (increases in SVR relative to cardiac output) towards

action-based harm have demonstrated a tendency to reject harm in moral dilemmas (Cushman et al, 2012). However, a link between moral judgments and physiological aversion to carrying out simulated harmful acts, was not found in a recent study that used an alternative measure of sympathetic cardiovascular arousal; pre-ejection period, which represents the force of heart contractility (Parton & McGinley, 2019). In contrast, utilitarian tendencies but not deontological tendencies, have been associated with lower resting heart-rate variability - a proxy measure for vagal tone, which is believed to support neurovisceral integration during moral decision-making (Park et al., 2016). The authors suggested that a reduced ability to integrate cardiac signals into the decision-making processes promoted stronger utilitarian preferences, but this finding was also not replicated in a later study using a small set of moral dilemmas (Parton & McGinley, 2019).

Measures of physiological arousal

The autonomic nervous system (ANS) regulates and maintains bodily functions and states, including our response to stressors (Montano et al., 2012). The sympathetic (SNS) and parasympathetic nervous system (PNS) represent two distinct branches of the ANS. The SNS (when activated) prepares the body for stress e.g. by increasing heart rate and blood-flow to the muscles ('fight or flight'). Typically, the PNS has an opposing role to reduce physiological stress-responses, e.g. by decreasing heart rate and increasing digestive processes ('rest and digest'). Motivational states of 'challenge' and 'threat' are associated with specific physiological patterns, including cardiovascular and hormonal changes (Blascovich & Mendes, 2000). 'Challenge' states arise when we evaluate personal resources as exceeding environmental demands, whereas 'threat' states arise when we evaluate demands as exceeding resources.

As discussed, an aversion to action-based harm has been associated with a physiological threat state and activation of the sympathetic nervous system, indicated by a shortening of pre-ejection period (PEP; Parton & McGinley, 2019) or increases in systemic vascular resistance (SVR) (Cushman et al., 2012). SVR is the amount of resistance in the circulatory system that must be overcome for blood to flow through. In states of 'threat' our bodies release norepinephrine which tightens blood vessels and increases the amount of resistance in the circulatory system (Mendes et al., 2007). Blood-pressure is related to SVR, however, SVR actually determines arterial blood pressure physiologically (Klabunde, 2011). Blood pressure can increase as a result of increased SVR (related to the tightening of blood

vessels i.e. vasoconstriction), but also as a result of overall cardiac output, volume or viscosity of blood etc. Cardiac-output (volume of blood ejected from the heart per-beat) increases in 'challenge' and 'threat' states (Mendes et al., 2007), therefore measures of blood-pressure alone could be a misleading indicator of underlying 'challenge' and 'threat' states.

Another measure of physiological arousal related to activation of the sympathetic nervous system is galvanic skin-response (GSR). GSR is a measure of electrodermal activity and may also indicate an aversion to harmful actions in moral dilemmas (McDonald et al., 2017; Moretto et al., 2010). GSR activation appears unrelated to systemic vascular resistance (Furedy & Gagnon, 1969) suggesting different neural processes (Prout, 1967). GSR is also sensitive to viewing violent media (Carnagey et al., 2007), and highly demanding cognitive processing (Botvinick & Rosen, 2009). Therefore, it is possible that increases in GSR during harm-based moral dilemma tasks may provide a more generalised indication of sympathetic arousal associated with violence or task-demands.

Moral dilemmas: harm aversion versus outcome maximisation

Trolley (Thomson, 1985) and Footbridge (Foot, 1967) type dilemmas used in moral psychology research have historically conceptualised utilitarian and deontological judgments as opposite ends of a scale. However, these responses are not considered to be inversely proportional as someone could condone the harm of one person (making the 'utilitarian' choice) because they are motivated to maximise outcomes for the most people or because they have a reduced sense of harm aversion (Conway & Gawronski, 2013). Conway and Gawronski (2013) developed a battery of moral dilemmas using a process dissociation (PD) technique (Jacoby, 1991) which allows the calculation of harm-rejection and outcome-maximisation tendencies separately and offers a more accurate understanding of the socio-emotional processes driving moral judgments (e.g. Reynolds & Conway, 2018). Responses to both conventional 'incongruent' moral dilemmas (i.e. causing harm whilst maximising outcomes), and 'congruent' dilemmas (i.e. causing harm but *not* maximising outcomes) are assessed to calculate the two parameters (Conway & Gawronski, 2013).

This procedure is increasingly favoured by researchers in this field (e.g. Park et al., 2016) and can further our understanding of the discrete processes and reactions that underlie judgments in moral dilemmas. For example, Reynolds and Conway (2018) recently clarified the role of emotional processes associated with action-aversion and outcome aversion in

moral judgments of harm. Interestingly, those scoring highly on action-aversion, demonstrated reduced dispositions to maximise outcomes (scored lower on the utilitarian parameter), as well as increased inclinations to reject causing harm (scored higher on the deontological parameter). To the contrary, those reporting higher outcome-aversion demonstrated a stronger inclination to maximise overall outcomes, while also showing greater inclinations to reject harm. Therefore, harmful actions can be experienced as emotionally aversive independent of their outcomes, but negative affect can also be experienced when contemplating witnessing harm befalling others, independent of the nature of the actions that cause harm (Reynolds & Conway, 2018). These findings suggest that there may be an emotional component to both harm-aversion and outcome-maximisation decisions which can be influenced by our aversions to harmful actions and outcomes, which has not previously been recognised due to the limitations of traditional moral dilemma paradigms (Reynolds & Conway, 2018).

Exploring moral behaviour in virtual reality

Research using immersive virtual reality (VR) has consistently demonstrated that people typically respond to events, situations and other people as if they were real (Pan & Slater, 2011; Riva et al., 2016; Slater & Sanchez-Vives, 2016). Immersive VR therefore represents a promising methodology for studying moral psychology (e.g. Francis et al., 2016; Pan & Slater, 2011), which his historically relied upon abstract representations of moral scenarios using text-based paradigms, yielding low ecological validity. The most significant finding of VR moral dilemma research so far is that moral judgments made in response to text-based paradigms often contradict moral behaviours carried out in more immersive VR environments (e.g. Francis et al., 2016; Francis et al., 2017; Patil et al., 2014). Therefore, behavioural measures of morality used in VR dilemmas are valuable for understanding discrepancies between text-based moral judgments and more realistic moral actions (Teper et al., 2011; 2015). Often people will be morally hypocritical and endorse hypothetical moral behaviours for social desirability reasons or based on what they believe to be the right course of action, whereas in reality moral actions can be personally costly (Batson et al., 1999). It is also more difficult to focus on emotions during hypothetical decision-making situations (Wilson & Gilbert, 2003), and changes in physiology triggered by a moral conflict such as increases in heart rate or breathing, are also likely to be less pronounced, which may lead to

divergent responses between moral judgments and more realistic moral behaviour (Patil et al., 2014; Teper et al., 2015).

Increased contextual saliency has proven to be an important factor in processing the value of a moral action (Patil et al., 2014). Patil et al (2014) found that differences in the contextual saliency (i.e. immersive VR versus not) of moral dilemmas corresponded to differences in autonomic arousal and utilitarian responses. The authors explain their findings as differences between value-representation targets (Cushman, 2013). The contextual saliency of the VR presentation influenced participants to be more sensitive to the consequences of their action (i.e. witnessing the deaths of virtual people), and therefore their aversive emotional state motivated them to minimise their own distress by choosing the option where fewer deaths occur (utilitarian). Whereas the authors suggest text-based presentation results in a value representation of the actions (deontological) i.e. pushing someone off a bridge accrues a negative value, subsequently resulting in fewer utilitarian responses. Patil et al (2014) suggest that the outcome-based value representation for not saving more people from harm is greater than the negative value associated with harming one person. Interestingly, post-hoc analyses indicated that the foregrounding of the virtual people was more important than watching their unpleasant deaths when making a utilitarian decision (Patil et al., 2014). In contrast, Navarette et al (2012) found no difference between text-based moral judgments and moral behaviour in contextually salient VR simulations of the trolley dilemma, however behavioural responses in VR were compared to text-based judgments from previous studies which could explain their results (see Patil et al., 2014).

A more recent study investigating personal force in a VR 'footbridge' simulation aimed to increase contextual saliency to a more extreme degree. Francis et al (2017) used a robotic manipulandum and interactive sculpture providing realistic haptic feedback to provide sensorimotor aspects relating to embodiment and touch. Francis et al (2017) used two experiments to compare differences in moral judgment and moral action. In experiment one, participants did not see the consequences of their actions in VR but received haptic feedback when carrying out a pushing action on a robotic manipulandum. Here, contextual saliency was absent as there was no visual element. In experiment 2, participants received haptic feedback when pushing a realistic sculpture, while also experiencing the moral dilemma in an immersive, contextually salient VR environment. The frequency of utilitarian moral actions was compared to the frequency of utilitarian moral judgments using the traditional text-based

paradigm. They found participants provided more utilitarian action responses in experiment 1 and 2, compared to moral judgments of text-based moral dilemmas. As contextual saliency was absent from experiment 1, the authors propose both a contextual salience and frame of reference explanation may account for the judgment-action discrepancy observed. For experiment 1, the authors suggest utilitarian choices were facilitated by an egocentric perspective generated by personal action choices (pushing the manipulandum device), as participants assessed the potential self-relevant consequences of condoning harmful action. In contrast, they proposed that moral judgments that asked about the moral acceptability of harmful acts, were more allocentric in nature and did not facilitate this perspective (see Tassy et al., 2013). Whereas, in experiment 2, the contextual saliency of the VR dilemma may have encouraged people to place more negative emphasis on the outcome of seeing the victims die (compared to non-contextually salient dilemmas) rather than the harmful act, as suggested by Patil et al (2014).

Despite these advances in VR methodologies in moral psychology, research has so far been limited to the traditional moral dilemma scenarios. Therefore, the process dissociation technique developed by Conway and Gawronski (2013) that allows independent assessment of harm aversion and outcome aversion tendencies has yet to be applied to an immersive virtual environment. A replica of the 20-dilemma battery developed by Conway and Gawronski (2013) is unlikely to be feasible in terms of both the resources required and the experimental-load participants would need to endure. However, a scaled version of this approach could be piloted in VR, whereby participants experience several matched 'incongruent' and 'congruent' moral dilemmas (Conway & Gawronski, 2013). This would allow some evaluation of their tendency to cause harm to others, in the absence of maximising outcomes.

Moral behaviour research in applied contexts

A critical caveat of traditional Trolley and Footbridge dilemmas, in addition to those already discussed, is that many people are now familiar with these problems, which could influence participants to respond in biased and unnatural ways (Pan & Slater, 2011). In addition, these problems bear little resemblance to real-life moral dilemmas we might typically face in real life. Highly controlled moral dilemma experiments that provide very clear and discrete moral choices typically using forced-choice paradigms, are beneficial for understanding how contextual, emotional or individual differences factors can influence

moral judgment and behaviour under experimental conditions (e.g. Patil et al., 2014). However, many have criticised these thought experiments as being too removed from real life to be useful to understanding real-world morality (Nyholm & Smids, 2016). For example, the uncertainty about the consequences of our actions is typically absent in moral dilemmas where we are provided with unrealistic levels of certainty about the outcomes of certain moral choices (Hansson, 2009; Pan & Slater, 2011).

Although some studies are attempting to make the sensorimotor experiences of traditional moral dilemmas more physically accurate (e.g. Francis et al., 2017), thought experiments such as the trolley and footbridge problems were conceived as "*stylized thought experiments we use for certain circumscribed purely theoretical and abstract purposes*" (pg.1285, Nyholm & Smids, 2016). These thought experiments are not necessarily meant to be applied to real-world moral behaviour. A much greater number of contextual variables and motivational drives are at play in real-life moral dilemmas, which if included in moral dilemma studies would decrease experimental control. However, some researchers (Pan & Slater, 2011; Faulhaber et al., 2018) are beginning to test 'applied' versions of moral dilemma scenarios in VR simulations to increase understanding of the factors influencing behaviour in more complex and realistic ethical dilemmas.

Pan and Slater (2011) tested an applied version of the trolley problem (n=36) in an immersive VR moral dilemma simulation to provide a novel and more plausible problem. In the scenario, an attacker is firing a gun at five visitors in an art gallery. The participant is in a virtual lift and they have the choice to bring the lift down to the ground floor, which would save five people on the top floor but put the person on the ground floor in danger. They found that participants' utilitarian responses in VR were consistent with moral judgment data collected prior to the study (n=80). Interestingly, in the VR simulation participants tended to panic and make mistakes in their immediate action (Pan & Slater, 2011). The authors suggested this may be because the choices for action were unclear, or because this is what people often do in real-world dilemmas; they freeze, make mistakes, and are conflicted about what to do. Faulhaber et al (2017) used a simulated moral dilemma VR driving scenario (n = 189). As participants were driving along a virtual road, a range of obstacles was placed in their path and they had to decide between two options to collide with. 'Obstacles' included a range of human avatars of different sizes and ages. Overall, they found participants responded in a utilitarian way - this was also the case in situations where participants had to

sacrifice their own avatar to save others. Interestingly, they found participants consistently saved younger avatars. However, it was not necessarily clear that hitting avatars would lead to death and there is a possibility that utilitarian actions in the first dilemma may have led to consistency in later dilemmas.

VR is gaining popularity as a method to explore driving behaviour in moral dilemma scenarios, due to the development of automated vehicles and the debate over the 'morality' algorithms these vehicles may eventually be controlled by. However, moral philosophers highlight important distinctions between the moral dilemmas like the Trolley problem and accident-algorithms of driverless cars (Nyholm & Smids, 2016). Accident algorithms are decisions made by multiple stakeholders prospectively, with an unlimited number of considerations and situational features, relying on risk estimation and decision-making under uncertainty, and must fundamentally account for moral and legal responsibility - issues which are still heavily debated. Whereas, in Trolley-type dilemmas; decisions are made in the present, only a small number of situational or human factors are considered, legal responsibility is removed, and facts and consequences are certain (Nyholm & Smids, 2016). Nevertheless, realistic VR driving simulations could be useful to understand what human factors (e.g. arousal, interoception) are important for moral behaviour in more contextualised and ecologically valid scenarios.

Rationale for proposed thesis research

Visceral sensations associated with changes in physiological arousal provide an adaptive function for motivating behaviour to support homeostatic regulation and efficient utilisation of energy (Craig, 2015; Seth & Friston, 2016). Our perception of somatic markers associated with changes in physiological arousal appear to be centrally important for shaping our emotional experiences, decision-making and behaviour (e.g. Craig, 2015; Damasio, 1996; Dunn et al., 2010; Farb et al., 2015). Negative affective arousal appears to provide a special regulatory function for guiding decision-making in harm-based moral dilemmas (Greene et al, 2001; 2004; 2009; Moretto et al., 2010; Reynolds & Conway, 2018). For example, physiological aversion to action-based harm has been shown to occur before conscious consideration of the consequences of harmful action (Cushman et al., 2012; Parton & McGinley, 2019). Little is currently known about how individual differences in interoceptive capacities correspond with the anticipatory physiological responses prior to moral decision-making/action, and the moral decision/action itself. Evidence thus far suggests both our

interpretations of internal signals and our ability to objectively perceive them are central to understanding how interoceptive information could influence our behaviour (Feldman Barrett et al, 2004; Garfinkel et al., 2015; Gu et al., 2013; Farb et al., 2015; Kever et al., 2015; Lazarus & Folkman, 1984).

Contextually salient immersive virtual reality (VR) simulations that can simulate sensorimotor processes associated with harmful action during a moral dilemma scenario, may offer a more ecologically valid setting to explore our emotional aversion to harm compared to traditional text-based paradigms (Francis et al., 2017; Pan & Slater, 2011). VR simulations of moral dilemmas are also more likely to generate stronger physiological and emotional reactions (Patil et al., 2014; Slater & Sanchez-Vives, 2016; Teper et al., 2015), as well as facilitate an egocentric perspective that is more likely to increase the saliency of self-relevant consequences of moral behaviour (Francis et al., 2017).

Understanding how subtle interoceptive processes are associated with physiological and emotional reactions to moral dilemmas, could have important implications for real-world moral behaviour and professional training programs (Francis et al., 2017; Grinberg & Hristova, 2013). For example, understanding the behavioural and physiological responses of drivers in anticipated collisions could inform the design of driver-control algorithms in automated vehicles that incorporate driver-data. Moreover, within professions such as the army or nursing, trained professionals are frequently confronted with moral dilemmas which have implications for the wellbeing of others (Corley, 2002). Interestingly, emergency services staff have shown reduced physiological arousal during moral dilemma tasks (Francis et al., 2017), which suggests that physiological reactivity in response to moral dilemma situations may need to be adapted in particular occupational contexts.

This PhD thesis investigates the role of interoception in moral decision-making and moral behaviour across four studies. Study 1 investigates whether interoceptive states of hunger, thirst, emotional state and interoceptive sensibility influences moral judgments of harm in an online cross-sectional design. Study 2 implements a range of physiological and interoception measures in an experimental lab study, to explore the link between physiological arousal, interoception and moral judgments of harm. Study 3 investigates the relationship between physiological arousal, moral behaviour and interoception in immersive VR moral dilemma scenarios in an autonomous vehicle. Finally, in light of the current pandemic, Study 4 explores the effects of harm-salience in news articles about coronavirus

and interoceptive sensibility, on real-world behavioural intentions during the pandemic and moral judgments regarding the treatment of coronavirus patients.

Study 1 Preface

Paper:

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). Hunger Bias or Gut Instinct ? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making. *Frontiers in Psychology: Perception Science*, *11*(September), 1–16. https://doi.org/10.3389/fpsyg.2020.02261

Pre-registration:

Brown, H., Fraser, D. S., and Proulx, M. J. (2019). *Does felt hunger and interoceptive sensibility influence moral judgments of harm?* Open Science Framework. osf.io/37vsh

Study 1 aimed to explore the influence of the everyday interoceptive states hunger and thirst on moral judgments of harm in people's naturalistic eating settings in a large online sample that could be recruited relatively quickly. This research was initially motivated by the protocol of some lab-based tasks that measure interoceptive accuracy and sensitivity including heartbeat counting tasks (Schandry, 1981), and water-drinking tasks (Van Dyck et al., 2016) used in Study 2 and 3. These tasks typically require people to abstain from eating and drinking for several hours to control for differences in for example caffeine or sugar consumption that may confound these tasks by altering people's momentary ability to detect internal states. However, short-term fasting manipulations have been shown to influence moral judgments of ethical violations (e.g. Vicario et al., 2018), changes in blood glucose have been associated with reduced levels of self-control and prosocial intentions (Gailliot et al., 2007), and fasting has been shown to increase people's awareness of changes in cardiovascular arousal (Herbert et al., 2012).

Interoceptive states of hunger and thirst indicate energy and fluid intake is deviating away from an expected range required to maintain homeostasis (Craig, 2015) which may be associated with individual differences in cognitive and emotional states (MacCormack, 2016; Stevenson et al., 2015). As negative emotional responses and increases in sympathetic physiological arousal are associated with harm-rejection judgments in hypothetical moral

dilemmas (e.g. Cushman et al., 2012; Greene et al., 2001; Moretto et al., 2010; Reynolds & Conway, 2018), we were interested in how hunger and thirst, a habitual tendency to focus on or notice bodily sensations (interoceptive sensibility; IS; Garfinkel & Critchley, 2013), and self-reported emotion and anxiety may interact to influence moral judgments of harm. Careful consideration of eligibility requirements ensured no important baseline differences in people's tendency to perceive internal sensations of hunger and thirst e.g. thyroid problems, gastrointestinal/heart conditions.

There were numerous novel elements in this study that contributed to the existing literature. No prior studies investigating the link between hunger and moral decision-making included measures of interoceptive sensibility (IS). However, Schnall et al (2008) found that IS was important in the link between disgust sensations and judgments of ethical violations, showing how individual differences in how we typically attend to our bodies can shape embodied moral judgment. In addition, we were specifically looking at the relationship between hunger, thirst, interoceptive sensibility and egocentric and allocentric moral judgments of harm, whereas prior studies have focused more on allocentric judgments of harmful acts (e.g. Vicario et al., 2018) and none have included measures of thirst, despite the close relationship with hunger (Mckiernan et al., 2008).

Furthermore, the moral dilemma stimuli used makes it possible to calculate independent parameters of harm-aversion and outcome-maximisation tendencies and is now considered a superior method to understanding the socio-emotional motivations behind these harm-based judgments, compared to earlier traditional dilemmas (Conway & Gawronski, 2013) such as the Trolley (Thomson, 1985) and Footbridge (Foot, 2003) problems. Conway and Gawronski (2013) drew attention to the importance of not conflating overt moral judgments with moral inclinations driving these judgments, which led them to develop a measure that could capture the relative strength of an inclination to avoid harm or maximise outcomes, which are not inversely related. In a series of studies exploring individual differences associated with these two motivations, Conway and Gawronski (2013) found that differences in harm-aversion tendencies were associated with empathy and perspective taking, whereas differences in outcome-maximisation tendencies were predicted by need for cognition. Therefore, these parameters show consistency with Greene's dual process model (Greene et al., 2001, 2004), supporting the notion that 'deontological' judgments are

associated with emotional reactions to the prospect of harm, and 'utilitarian' judgments are associated with more cost-benefit deliberations concerning the outcome of harmful actions.

The 'Utilitarian' (outcome-maximisation) and 'Deontological' (harm-aversion) parameters are calculated with the following formulae (Conway & Gawronski, 2013), where 'action' refers to condoning harmful action and 'inaction' refers to rejecting harmful action:

Percentage (%) inaction incongruent trials = 1 - % action on incongruent trials
Percentage (%) inaction congruent trials = 1 - % action on congruent trials
Utilitarian parameter = % inaction congruent trials – % inaction on incongruent trials
Deontological parameter = % inaction congruent trials/ (1-Utilitarian parameter)

We discriminated between egocentric and allocentric moral judgments (Tassy et al., 2013) as egocentric judgments have been shown to engage emotional processes in the brain that allocentric judgments do not (Berthoz et al., 2006). We expected egocentric moral judgment would be more likely to be associated with interoceptive processes, as both involve a level of self-referential processing. In addition, we included a measure of intuitive decision-making, the Cognitive Reflection Task (Frederick, 2005) to explore whether people's tendency to focus or notice internal sensations predicted people's ability to disregard 'gut' responses to counter-intuitive problems, to explore whether this was a potential mechanism through which interoceptive sensibility may influence moral decision-making. Finally, we discussed correlations with possible confounding factors such as age, sex and anxiety to explicate potential interrelationships among the independent and dependent variables to contribute to the wider individual differences and moral judgment literature.

Note: Power analysis

As there are currently no studies showing the effect size for interoception on moral judgments, the power analysis was based on a conservative estimate of a small to medium effect size (f^2= 0.10) of interoception in regression analyses. The degrees of freedom (*df*=8) were chosen to account for the potential of conducting moderation, or moderated mediation analyses with several predictors e.g. between interoception, emotion, hunger, thirst etc.- although this was not necessary in the end.

Statement of Authorship

This declaration concerns the article entitled:			
Hunger Bias or Gut Instinct? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making			
Publication status (tick one) **Draft manuscript** ☐ **Submitted** ☐ **In review** ☐ **Accepted** ☐ **Published** X			
Publication details (reference)	Brown, Helen, Proulx, M. J., & Fraser, D. S. (2020). Hunger Bias or Gut Instinct ? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making. *Frontiers in Psychology: Perception Science, 11*(September), 1–16. https://doi.org/10.3389/fpsyg.2020.02261		
Copyright status (tick the appropriate statement) I hold the copyright for this material: X Copyright is retained by the publisher, but I have been given permission to replicate the material here: ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed the formulation of ideas, design of methodology, experimental work and presentation of data in journal format. The candidate (H.B.) conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	This information has been redacted for privacy reasons.	**Date**	05/11/2020

Hunger-bias or gut-instinct? Responses to judgments of harm depending on visceral state versus intuitive decision-making

Helen Brown[1, 2], Michael J. Proulx [2], Danaë Stanton Fraser[1]

[1] CREATE Lab, University of Bath, Psychology Department, Bath, United Kingdom

[2] Crossmodal Cognition Lab, University of Bath, Psychology Department, Bath, United Kingdom

Abstract

Empirical investigation into the emotional and physiological processes that shape moral decision-making is vast and growing. Yet, relatively less attention has been paid to measures of interoception in morality research despite its centrality in both emotional and physiological processes. Hunger and thirst represent two everyday interoceptive states, and hunger, in particular, has been shown to be influential for moral decision-making in numerous studies. It is possible that a tendency to focus on internal sensations (interoceptive sensibility), as well as the emotional and physiological states associated with visceral states, could be important in the relationships between hunger, thirst, and moral judgments. This cross-sectional online research (n=154) explored whether interoceptive sensibility, hunger, thirst and emotional state influenced appropriateness and acceptability judgments of harm. The moral dilemma stimuli used allowed the independent calculation of 1) people's tendency to avoid harmful action at all costs and 2) people's tendency to maximise outcomes that benefit the greater good. The Cognitive Reflection Task (CRT) was implemented to determine whether an ability to override intuitive responses to counterintuitive problems predicted harm-based moral judgments, as found previously. Hunger-bias, independent of interoceptive sensibility and emotional state was influential for non-profitable acceptability judgments of harmful actions. Contrary to dual-process perspectives, a novel finding was that more intuitive responses on the CRT predicted a reduced aversion to harmful actions which was indirectly associated with interoceptive sensibility. We suggest that interoceptive sensibility may indicate people's vulnerability to cognitive miserliness on the CRT task and reduced deliberation of moral dilemma stimuli. The framing of moral dilemmatic questions to encourage allocentric (acceptability questions) versus egocentric perspectives (appropriateness questions), could explain the divergence between hunger-bias and intuitive decision-making for predicting these judgments respectively. The findings are discussed in relation to dual-process accounts of harm-based moral judgments and evidence linking visceral experiences to harm-aversion and moral decision-making.

Keywords: interoception, moral judgment, hunger, decision-making, moral dilemmas.

Introduction

Our current homeostatic needs provide a context for decision-making (Craig 2015; Gailliot, 2013; Yam et al., 2014). Important decisions sometimes with serious consequences such as prescribing anti-biotics (Linder et al., 2014), judicial rulings (Danziger et al., 2011) and voting behaviour (Gomez et al., 2007) can be influenced by regularly occurring trivialities such as the time of day (Linder et al., 2014; Danziger et al., 2011), bad weather (Gomez et al., 2007), carbon-dioxide levels (Satish et al., 2012) and how hungry we are (Gailliot, 2013). The connection between how we feel right now and the decisions we make is no coincidence. Interoception refers to our perception and interpretation of visceral sensations associated with homeostatic regulation inside the body, such as those originating in the cardiovascular, respiratory, and gastro-intestinal systems (Craig, 2015; Garfinkel & Critchley, 2013). Brain areas responsible for the perception of visceral states (e.g. feeling hot, cold, full) are also implicated in the integration of this information to initiate drive states (e.g. hunger, thirst, sex-drive) that in turn affect how we feel (Craig, 2015). The vagus nerve communicates the majority of information from visceral centres to the brain stem (Hellström & Näslund, 2001) coordinating adaptive fight/flight responses on one hand and emotional expression and social engagement processes on the other depending on the physiological state of the body (Porges, 2007). There is considerable crossover in brain areas responsible for interoception, emotion and social cognition (Adolfi et al., 2017) and empirical advances in the field of embodied cognition continue to illuminate how cognitive products of the mind can be rooted within the body (Häfner, 2013). Furthermore, individual differences in how we perceive internal sensations have been shown to be important in the link between visceral processes and decision-making (Dunn et al., 2010; Häfner, 2013).

Visceral states like hunger can influence ethical decisions in the lab (e.g. Yam et al., 2014) and the real-world (e.g. Gailliot, 2013). Hunger is the subjective experience of food deprivation comprising visceral sensations in the stomach area, an emotional desire or wanting to eat, and cognitive states associated with eating, food, and hunger (Stevenson et al., 2015). Thirst is a comparatively understudied but related drive, largely regulated by food intake (Mckiernan et al., 2008) and comprises a desire or wanting to obtain and drink water, often accompanied by sensations such as dryness of mouth (Ramsay & Booth, 2012). Incidental emotional states can influence moral decision-making (Valdesolo & Desteno, 2006), sensitivity to moral norms (Gawronski et al., 2018) and emotional-regulation difficulties predict a bias towards immoral judgments (Zhang et al., 2017). Differences in

blood glucose levels have also shown to predict prosocial intentions (Gailliot et al., 2007). Danziger and colleagues found the probability of court judges to provide less favourable rulings was increasingly likely before the provision of a food/rest break compared to afterwards (Danziger et al., 2011). However, other researchers (Weinshall-Margel & Shapard, 2011) have contested this, suggesting that the order of cases seen by judges was partly responsible for this observation.

Lab-based research has been more effective at substantiating a link between hunger and moral judgments as hunger can be objectively manipulated. Vicario et al (2018) found hunger reduced moral disapproval ratings for ethical violations, suggesting hunger-bias may reduce the harshness of moral judgments. A dispositional sensitivity towards feelings of disgust was also found to increase the severity of moral disapproval ratings of ethical violations. Vicario and colleagues suggested hormonal reactions and interoceptive signals triggered by eating may evoke feelings of nausea interpreted as disgust (Tracy et al., 2019) which subsequently inform moral judgments. This is consistent with other work (Horberg et al., 2009; Wheatley & Haidt, 2005), including Schnall et al (2008) who found disgust manipulations encourage harsher judgments of ethical violations and is strongest for those with a greater tendency to pay attention to interoceptive sensations. Despite large variations in interoceptive sensitivities between people and daily fluctuations in interoceptive states, individual differences in interoception is an underexplored area in the link between moral decision-making and visceral states like hunger (Dunn et al., 2006).

Damasio's (1996) somatic marker hypothesis (SMH) was among the first theoretical frameworks to reveal the neuropsychological foundations that connect fundamental visceral processes with higher-level moral cognitions. The SMH (Damasio, 1996) describes how changes in bodily states have the potential to alter our emotional state and bias our thinking processes to support adaptive behavioural responses to the environment (Craig 2015; Barrett, 2016). The ventromedial prefrontal cortex is believed to be responsible for the representation of homeostatic information (including emotional state) when evaluating ethical violations (Damasio, 1994; Moretto et al., 2010). Damage to this area is associated with emotional deficits in guilt and empathy (Anderson et al., 2013), reduced physiological responses to moral decisions and greater acceptance of moral violations (Moretto et al., 2010). The insula is a key centre for interoceptive integration (Adolfi et al., 2017) and is implicated in processing negative emotional states, particularly disgust sensitivity (Calder et al., 2007)

which can bias moral decision-making (Greene et al., 2004). It is possible people with superior ability to perceive interoceptive processes, could be more influenced by this information when forming moral judgments. For example, in research using the Iowa Gambling Task (a card-choosing task measuring decision-making under uncertainty), people with a superior ability to detect internal sensations were more influenced by concurrent somatic signals even when those signals unhelpfully guided them towards high-risk card-decks (Dunn et al., 2010).

Historically, research exploring emotional influences in moral decision-making have focused on harm-based moral dilemmas like the Trolley (Thomson, 1985) and Footbridge (Foot, 2003) problems, as particularly emotive moral-conflicts to consider (Greene et al., 2001). In these dilemmas' participants judge whether it is acceptable to cause fatal harm to one person either directly (Footbridge) or indirectly (Trolley), as a necessary means to saving the lives of more (>1) people. Judgments can be influenced by an emotional reaction to the harmful *action* towards the one person intentionally harmed ('deontology'), or to the *outcomes* of the action for the many people who would be harmed otherwise ('utilitarianism') (Cushman, 2013; Miller et al., 2014). This traditional moral dilemma paradigm places utilitarianism and deontology on opposite ends of a bi-polar scale, preventing us from determining whether someone chooses to harm the 1 person because they have a weakened aversion to harming others or because they are more motivated to save the lives of the most people (Conway & Gawronski, 2013). A more recent process-dissociation approach (Conway & Gawronski, 2013) uses moral dilemma stimuli that allow the measurement of people's outcome-maximisation (utilitarian) and harm-aversion (deontological) motivations independently. This method works by calculating the probability that someone chooses to condone harming others when harm results in a 'greater good' overall, and when it does not. Although people's tendencies to avoid harm or maximise outcomes do not necessarily represent people's abstract views about deontological and utilitarian philosophies (Kahane et al., 2018), these terms are used for clarity.

Deontological moral judgments associated with the rejection of harmful action, have been associated with more visceral and intuitive decision-making processes than utilitarian decisions (Greene et al., 2001; Park et al., 2016). Greene's (2001) dual-process account of morality proposes deontological judgments are driven by automatic and emotional responses associated with activation of emotional centres in the brain, whereas, utilitarian judgments are

driven by more reflective, cognitive processes and are associated with activation of brain areas implicated in cognitive control (Greene et al., 2004). In support of a dual process conceptualisation, emotional arousal predicts deontological preferences (Szekely & Miu, 2015) and performing or witnessing harmful actions correlates with measures of cardiac arousal (Cushman et al., 2012; Parton & McGinley, 2019). More calculative reasoning styles have been associated with utilitarian response tendencies (Patil et al., 2020) and successful performance on the Cognitive Reflection Task (CRT; Frederick, 2005) is associated with increased utilitarian judgments, potentially due to its' association with cognitive deliberation (Baron et al., 2015). The CRT task includes questions that have both correct and 'intuitive' answers and can be scored according to correct versus intuitive responses (Erceg & Bubić, 2017). Successful performance on this task requires some reflection to avoid the intuitive lures and determine the correct solutions. As such, this task is believed to provide an indication of a person's ability to 'override' their gut response to counter-intuitive problems (Frederick, 2005). Byrd and Conway (2019) suggest that arithmetic-reflection ability (captured by the CRT) is responsible for the association with utilitarian preferences, possibly because it indicates a greater numerical focus (i.e. saving *more* lives) when weighing up moral decisions. Whereas, Park and colleagues (2016) suggest strong utilitarian preferences may reflect poorer integration of visceral signals into the decision-making process, leading participants to place more weight on the outcomes of harmful action.

The physiological, emotional and cognitive processes implicated in moral decision-making are relevant to consider in the context of hunger and thirst, as changes in our psychophysiological states have the potential to bias decision-making processes (Critchley & Garfinkel, 2018). Food deprivation is often associated with increased physiological arousal (e.g. Chan et al., 2007; Ribeiro et al., 2009). Ghrelin (the 'hunger' hormone) appears to play a role in regulating our responses to stressors potentially by increasing anxiety (see Korbonits et al., 2004) and relationship with the stress hormone cortisol (Sarker et al., 2013). Although there has been less empirical interest in thirst, available evidence suggests hydration levels do not affect cardiovascular reactivity (Schwabe et al., 2007) but can affect blood-reactivity to stress (Rochette & Patterson, 2005). Cardiovascular arousal is of particular interest, as arousal represents a core component of emotional experience (Barrett & Russell, 1999) which can intensify the processing of emotionally salient information (McGaugh, 2015) and could influence moral decision-making (Greene et al., 2001). Heartbeat signals alone can directly influence cognition and facilitate the detection of fearful and threatening stimuli (Garfinkel &

Critchley, 2016). In addition, the sound of 'quickening' heartbeat feedback has shown to predict moral decision-making (Gu et al., 2013), demonstrating how even a belief that we are physiologically aroused can influence our moral choices. Hunger sensations or sensations associated with hunger-induced physiological arousal may manifest as different psychological states (MacCormack, 2016; Barrett et al., 2004) depending on individual differences in perception (Dunn et al., 2010; Herbert et al., 2012) and interpretation (Domschke et al., 2010) of these interoceptive processes. For example, brain regions associated with the conscious awareness of interoceptive states are also implicated in subjective emotional experience (Zaki et al., 2012) and individuals who are better at detecting heartbeat sensations experience more arousal-focused emotional experiences (Barrett et al., 2004). Furthermore, preliminary evidence suggests hunger could actually provide a context for more accurate perception of visceral sensations due to changes in the autonomic nervous system that alter cardiac activity (Herbert et al., 2012). Therefore, although subjective hunger and thirst states may be influential for moral decision-making due to the physiological experiences typically accompanying them, it is likely that individual differences in interoceptive sensitivities will shape how these visceral states translate into psychological and emotional states.

Interoceptive sensibility (IS) is one construct that could influence the psychological manifestation of visceral states and is a measure of a person's tendency to focus on internal sensations, independent from their ability to objectively detect internal sensations (Garfinkel & Critchley, 2013). Although some evidence suggests heartbeat detection accuracy corresponds with increased sensitivity to bodily information (Duschek et al., 2015) other research indicates interoceptive accuracy and sensibility are unrelated (Ainley & Tsakiris, 2013; Ferentzi et al., 2018). Individual differences in IS has shown to be important in the link between our visceral experiences and subjective appraisals of these experiences (Häfner, 2013) and could potentially shape the interpretation of visceral sensations present during moral decision-making. Individuals high in body awareness typically direct more attention towards visceral sensations, increasing the likelihood they will observe and misinterpret physiological changes as meaningful which can influence emotional state (Palomba & Stegagno, 1995) and increase anxiety (Clark et al., 1997; Domschke et al., 2010). Paulus and Stein (2010) suggest that visceral sensations detected by people with high levels of anxiety can be intensified and associated with bad or aversive outcomes and is consistent with the finding that IS can increase risk-averse behaviour when bodily information is present

(Salvato et al., 2019). Overall the link between anxiety and moral judgments of harm presents a mixed picture. Anxiety facilitates increased vigilance to threats and has been associated with unethical behaviour (Kouchaki & Desai, 2015). There is some evidence to suggest that self-oriented anxiety associated with empathy, can increase people's tendency to reject harm in traditional moral dilemmas (Sarlo et al., 2014). Trait anxiety has shown to specifically predict moral goodness ratings of utilitarian action in the Footbridge dilemma, whereas mild anxiety-inducing manipulations appear to have less of an impact on moral judgments (Zhao et al., 2016). It is plausible that a greater attentional focus on bodily sensations could heighten sensitivity to arousal-based physiological sensations accompanying hunger or thirst which, if interpreted as meaningful and anxiety-evoking (Paulus & Stein, 2010) could influence moral decision-making (Sarlo et al., 2014; Zhao et al., 2016).

Importantly, prior studies exploring the relationship between hunger and moral judgments have measured judgments of ethical violations, which require people to make allocentric judgments about the acceptability of other people's morally dubious actions (e.g. Vicario et al., 2018). However, moral dilemmas used to explore people's aversion to harm typically ask questions that facilitate an egocentric perspective e.g. *'Would **you**, carry out X action…in order to?'* (e.g. Foot, 2003; Conway & Gawronski, 2013; Thomson, 1985). Several studies have found discrepancies between whether people judge another person's actions to be morally acceptable, versus whether people agree that they would perform 'immoral' actions themselves (Pletti et al., 2017; Tassy et al., 2013). An egocentric perspective, that is putting ourselves in the shoes of the agent committing an immoral act, encourages us to consider the self-relevant consequences of our actions (Sood & Forehand, 2005). Egocentric moral judgments, but not allocentric judgments have been associated with activation of the amygdala, suggesting these judgments rely on emotional processes that allocentric judgments do not (Berthoz et al., 2006). Therefore, it is possible that imagining ourselves personally performing harmful acts could influence how likely we are to refer to bodily and emotional cues when forming moral judgments. Extending previous work, we explored whether the role of hunger, interoceptive process and emotional state, were associated with moral appropriateness (egocentric) and moral acceptability (allocentric) judgments of harm in the same way. Furthermore, comparing people's tendency to judge harmful acts as morally acceptable from an allocentric perspective when harm results in a greater good, and when it does not, has not previously been explored.

We do not yet have a clear understanding of how incidental visceral and emotional states may interact and exert influence over moral judgments in the moment, as the relationships between these variables are complex and multi-directional. Food deprivation can affect physiological arousal (e.g. Chan et al., 2007; Korbonits et al., 2004) and emotional processes (MacCormack, 2016) which are known to influence moral judgments regarding the harm of others (Damasio et al., 1990; Greene et al., 2001; Cushman et al., 2012; Parton & McGinley, 2019). Hunger also influences interoceptive processes and may even heighten our awareness of changes in cardiac arousal (Herbert et al., 2012). A heightened awareness of internal sensations associated with hunger/thirst may increase the availability of bodily cues (Domschke et al., 2010). Hunger states could therefore influence moral decision-making e.g. by reducing the harshness of moral acceptability judgments (e.g. Vicario et al., 2018), but the direction of this effect has not previously been investigated with harm-based moral judgments. Emotional state is fundamentally linked with interoceptive processes, and hunger (Barrett, 2016; MacCormack, 2016; Macht & Simons, 2000) and can affect moral judgments (e.g. Valdesolo & Desteno, 2006; Zhang, Kong, & Li, 2017). People's current emotional experiences could therefore modulate the relationship between hunger/thirst and moral decision-making. We also explored the influence of sex, age and individual differences in anxiety for predicting moral judgments. Women and older people are more likely to reject harmful action in hypothetical moral dilemmas (McNair et al., 2019; Armstrong et al., 2019). Anxiety is associated with heightened cardiac arousal which can affect how we process threatening information (Garfinkel & Critchley, 2016) and is a psychological correlate of both hunger (Herman et al., 1987) and interoceptive sensibility (Domschke et al., 2010). The role of anxiety in moral decision-making appears mixed. Anxiety has shown to increase unethical behaviour in some circumstances (Kouchaki & Desai, 2015), with trait anxiety and self-focused emotional distress demonstrating varying influences on moral judgments (Sarlo et al., 2014).

The current study aimed to assess the interdependent relationships between interoceptive sensibility, hunger and moral judgments of harm with the following research questions. The protocol was registered on the Open Science Framework (OSF) (Brown et al., 2019).

R1. Does felt hunger or thirst bias responses to a moral judgment task?

R2. Does interoceptive sensibility moderate the relationship between hunger/thirst and moral judgments of harm?

R3. Does emotional state moderate the relationship between hunger and moral judgments of harm?

R4. Does sex, age and/or anxiety predict moral judgments of harm?

Materials and Methods

Design

This was a within-subjects cross-sectional study (n=154) testing pre-registered research questions and exploratory hypotheses in a series of regression analyses. Moral appropriateness and moral acceptability judgments were the dependent variables. Hunger, thirst, interoceptive sensibility, incidental emotional state, and performance on the Cognitive Reflection Task (Frederick, 2005) were the independent variables. The influence of age, sex and anxiety for predicting moral judgments was also explored.

Measures

Demographics

Participants completed a brief demographic form indicating their sex, age, nationality and ethnicity. Collecting sex data was preferred over gender, as physiological sex differences were more relevant due to known sex-differences in interoceptive abilities. Experience in mindfulness/mediation practice was collected as a control variable due to its associations with body awareness (Bornemann et al., 2015) which could inform interpretation of the results. The item read: '*Are you an experienced meditator or regularly practice mindfulness?'* with response options: *No/Practice mindfulness or mediate occasionally/Yes*, coded for analysis.

Health questionnaire

A brief health questionnaire was used to assess participants' general health on the day prior to, and day of the experiment in the interest of managing any outliers that could influence the dependent and independent variables e.g. feelings of nausea, sickness. Only one of the questions regarding 'current state of health' was coded for analysis as it was deemed more relevant to the participants current emotional state. The item read: *'How is your overall health at this moment?*', and response options included: *Very Bad/Unwell, Slightly Unwell, No Complaints, Fine, Very Good*. These were numerically coded 1 to 5 pre-analysis to create a measure of 'current health'.

Anxiety

State and trait anxiety were measured using the State and Trait Anxiety Scale (STAI, Spielberger & Gorsuch, 1983). This consists of two identical 20-item scales that ask participants to rate how they feel *right now* (State anxiety) and how they feel *in general* (Trait anxiety). Participants were asked to indicate their agreement (*Not at all/Somewhat/Moderately so/Very much so*), with twenty different statements e.g. '*I feel calm', 'I feel tense', 'I feel at ease'*. The scales include positive and negatively coded items to calculate two cumulative scores representing State and Trait anxiety.

Interoceptive sensibility

Interoceptive sensibility (IS) concerns individuals' beliefs about their sensitivity to normal bodily processes (Ferentzi et al., 2018; Garfinkel & Critchley, 2013) and was measured using the 'Private Body consciousness' subscale of The Body Consciousness Questionnaire (Miller et al., 1981). This subscale offers a parsimonious measure of interoceptive sensibility, focusing specifically on bodily sensations and is commonly used in interoception research (e.g. (Ainley & Tsakiris, 2013; Sze et al., 2010; Werner et al., 2009). The entire Body Consciousness Questionnaire (Miller et al., 1981) was used in the interest of maintaining scale-validity. Only scores for the Private Body Consciousness subscale (PBCQ) were calculated for analysis which includes 5 questions measuring how often people typically notice or pay attention to interoceptive sensations. Subscale items include: *'I know immediately when my mouth or throat gets dry', 'I am sensitive to internal bodily tensions'* and '*I am quick to sense the hunger contractions in my stomach*'. Participants indicated how characteristic each statement was of themselves on a scale (extremely uncharacteristic/uncharacteristic/ neutral/characteristic/extremely characteristic*).* Items were

numerically coded 1-5 resulting in a maximum possible score of 25. Mean scores were calculated for all participants before analysis.

State emotion

State emotion was measured using the Positive and Negative Affect Scale (PANAS; (Watson et al.,1988). Positive (PA) and negative (NA) affect represent two independent subscales of subjective emotional experience. Each subscale consists of 10 items and demonstrate high internal reliability (PA: Cronbach α= .89; NA Cronbach α = .85; Crawford & Henry 2004).

Moral judgment stimuli

Conway and Gawronski's (2013) moral dilemma stimuli were used to assess independent response-tendencies of harm-aversion and outcome-maximisation separately. Harm-aversion represents the tendency to reject the moral appropriateness of carrying out harmful actions regardless of the consequences of harm. Whereas outcome-maximisation represents the tendency to condone the moral appropriateness of harmful action, if harm results in a greater good overall (e.g. saving more lives). Participants respond with a yes or no response (see Table 1). Their text-based procedure includes 20 moral dilemmas made up of 10 story-pairs, which are matched in story-content but the amount of 'greater good' resulting from a harmful action is manipulated. This makes it possible to measure harm-aversion tendencies while controlling for people's tendency to maximise outcomes. A harm-aversion and outcome-maximisation score is calculated for each person, based on the probability that they accepted/rejected harm when harm resulted in a 'greater good', and when it did not (for full probability equations see Conway & Gawronski, 2013).

Table 1.

Matched pair of moral dilemma stimuli from Conway and Gawronski's (2013) moral dilemma task

Causing harm maximises outcomes	**Causing harm does NOT maximise outcomes:**
You are a soldier guarding a border checkpoint between your nation and one troubled by insurgent violence. You notice a young man in a cheap car approaching the checkpoint with a determined look on his face. You suspect he means to bomb the checkpoint, **killing all the soldiers inside.** He is rapidly approaching your station. **Question:** Is it appropriate for you to shoot and kill the approaching man? Yes/No	You are a soldier guarding a border checkpoint between your nation and one troubled by insurgent violence. You notice a young man in a cheap car approaching the checkpoint with a determined look on his face. You suspect he means to cross the border in order to **work illegally inside your country**. He is rapidly approaching your station. **Question:** Is it appropriate for you to shoot and kill the approaching man? Yes/No

Moral acceptability ratings

Moral acceptability ratings of allocentric ethical violations (Schnall et al., 2008) or moral goodness ratings of harmful actions (Sarlo et al., 2014) provide a useful scale measure of the strength of people's judgments of harmful actions. Here we implemented a moral acceptability measure, to capture the strength of people's allocentric moral judgments when harmful action results in a greater good and when it does not. Following each of the moral dilemmas, we asked participants to judge the moral acceptability of the harmful actions proposed in the previous moral dilemma (see Table 1). The item read: *'How morally acceptable or morally unacceptable do you find the proposed action to be?'* Response options included: 1 = Completely unacceptable, 2 = Moderately unacceptable, 3 = Slightly

unacceptable, 4 = neither acceptable nor unacceptable, 5= Slightly acceptable, 6= Moderately acceptable, 7= Completely acceptable (adapted from Schnall et al., 2008).

Hunger and thirst

Two separate, single-item visual analogue scales were used to assess self-reported sensations of Hunger and Thirst on a scale of 1 to 9: *'How Hungry/Thirsty do you feel at this moment?'* (1=not at all, 9=extremely hungry/thirsty). Hunger and Thirst were assessed last to avoid any priming-effects before the moral judgment task.

Cognitive Reflection Task (CRT)

The original CRT (Frederick, 2005) assesses participant's ability to override intuitive or 'gut' responses to counter-intuitive problems. The task involves three questions that have both an intuitive and correct answer e.g. '*A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?*'. Participants manually typed their answers and response-time was not capped. Successful performance on this task requires further deliberation of the questions to determine the correct solutions and therefore, better performance is associated with a greater ability to override the 'intuitive' or more obvious answer. This measure aimed to capture participant's intuitive versus analytic decision-making tendencies when faced with counterintuitive problems. There are many possible scoring methods for the CRT. As such, both the 'Regular' scoring method (totaling only correct answers) and the 'Intuitive' scoring method (totaling only intuitive answers and disregarding incorrect answers) were used (Erceg & Bubić, 2017), to inspect correlations between these alternative calculations.

Procedure

Following approval from University of Bath Psychology Ethics Committee, 154 participants were recruited online via advertisements displayed on University of Bath research participation portal and social networking sites. The experiment was developed in Qualtrics and accessible via an anonymous web-link. All partially completed questionnaires were excluded from analysis. We exceeded our target sample size of 120 participants, which was based on a-priori power calculations using G*Power for multiple linear regression models, assuming $\alpha = .95$, $\beta = .8$ and $f^2 = 0.10$ (df=8). Inclusion criteria for participation was guided by a literature review of physiological and psychological confounds known to influence the primary independent variables, namely hunger, thirst and interoception.

Participants were required to be aged 18+; with no current mental health issues; no history of disordered eating, diabetes, thyroid conditions, gastrointestinal or heart conditions or previous surgery to those areas; and no current health conditions or medication that affected diet, weight or exercise. Eligibility criteria were emphasised on the research advertisements and participant information sheet.

Potential participants accessing the experiment in Qualtrics were first presented with a study information sheet. They were then asked to confirm they met eligibility requirements and encouraged to contact the experimenter with any questions or concerns about taking part. Participants then completed an online consent form and were made aware they could enter their names into a prize draw at the end of the experiment in exchange for their participation. Participants worked through a series of questionnaires in the order outlined below with instructions provided before each questionnaire. The experiment took roughly thirty minutes to complete and could be mostly carried out at the pace of the respondent. The moral dilemma task was the only timed element of the experiment, whereby the text for each moral dilemma story would time-out after 45 seconds and was followed by the moral judgment questions. Participants could advance to the questions after 20 seconds with a button click. This ensured reading time for each moral dilemma was roughly standardised and was clearly signposted in the instructions before starting the task. Upon completion of the study participants were thanked for their time and provided with some further information about the study and experimenter contact details. They were then asked if they would like to enter the prize draw to win 1 of 4 £25 Amazon vouchers, by entering their details via an anonymous link to a Raffle survey in Qualtrics.

Results

Data reduction and descriptive analysis

The sample was 31.8% male and the age of participants ranged between 18-70 years (Median=31, SD=12.21). Statistical analysis was carried out with SPSS v.24. A Pearson's bivariate correlation analysis including all of the variables was conducted first, followed by a series of ordinary least squares regression analyses to address preregistered and exploratory hypotheses. The SPSS scripts for moderation, mediation and conditional process analyses (PROCESS) were adopted from Hayes (2018). For all moderation analyses carried out in PROCESS, interactions are probed at the 16^{th}, 50^{th} and 84^{th} percentiles by default. As females were considerably overrepresented in this sample, a bootstrapping method was adopted for all

regression analyses (5000 x bootstrapping samples, 95% confidence interval) to generate standard error estimates that do not rely on parametric assumptions (Hayes, 2018).

Dependent variables: moral judgments and moral acceptability ratings

The four moral judgment dependent variables included 1) harm-aversion and 2) outcome maximisation tendencies and moral acceptability ratings for 3) congruent and 4) incongruent trials. Raw harm-aversion and outcome-maximisation scores were standardised into z-scores as suggested (see supplementary material; Conway and Gawronski 2013). As expected, harm-aversion and outcome-maximisation scores only showed weak negative correlation (r = -.092, p=.259); confirming the independence of these response tendencies. To explore whether people judged harmful actions (from an allocentric perspective) as more morally acceptable for trials where harm maximised outcomes and when it did not, moral acceptability ratings for the harmful actions proposed in each moral dilemma were averaged for trials where harm did not maximise outcomes (congruent), and trials where harm maximised outcomes (incongruent) (see Table 1). This resulted in two average moral acceptability scores for each participant: 1) acceptability_incongruent and 2) acceptability_congruent. Each acceptability score represented 10 moral acceptability ratings. Moral acceptability scores for congruent and incongruent trials were strongly positively correlated (r=.640, p<.001). This indicates people were relatively consistent in how morally acceptable they judged harmful actions to be from an allocentric perspective across all trials, when harm maximised outcomes and when it did not.

The distribution of studentised residuals of the dependent variables were inspected. Outcome-maximisation scores and acceptability_incongruent scores were fairly normally distributed. Harm-aversion scores sat slightly higher than the mean on average, however only mild skewness was identified. A log10 transformation was carried out on acceptability_congruent scores to adjust for a strong positive skew. For all regression analyses, a casewise diagnostic was performed on studentised residuals to identify outliers affecting the values of the estimated regression coefficients. Only three outliers ±3 standard deviations were identified overall and removed from the associated regression analysis. A Cook's distance and Levene's test confirmed no leverage values or unusual data points in each regression model. All other regression assumptions were met.

Scale reliability

The state emotion and trait and state anxiety measures showed high internal reliability. Coefficient *a* was .87 for positive affect, .89 for negative affect (.82 for the entire PANAS measure), .94 for state anxiety and .95 for trait anxiety. Adequate internal consistency was found for the private body consciousness subscale (5 items) of the Body Consciousness Questionnaire (a=.65) and is comparable to prior research (Christensen et al., 1996). Scores for the Cognitive Reflection Task (CRT) were coded for the presence of correct answers (Regular scoring) and intuitive answers (Intuitive scoring) and demonstrated very high negative correlation (r=-.910, p<.001). Correlations for the independent variables can be found below in Table 2.

Hunger and thirst

The mean hunger rating was 3.24 (SD=2.09) and 4.07 for thirst (SD=1.96), and the mode being 1 for hunger and 3 for thirst. As expected, hunger and thirst were positively correlated (r=0.294) and both scores positively predicted how many hours it had been since participants reported eating. Hunger and thirst scores were relatively normally distributed although participants typically reported less felt hunger than the median response option. Thirst was positively correlated with interoception (r=.247, p<.001) suggesting that people who were more likely to focus on internal sensations were also more likely to report subjective experiences of thirst.

Anxiety, emotion and interoceptive sensibility

Both state and trait anxiety strongly positively correlated with negative affect and were negatively correlated with positive affect (Table 2), which is expected as subjective arousal comprises a core component of affective experience (Barrett & Russell, 1999). More anxious people were more likely to report feeling unwell and although the direction of the relationship is unclear, correlation between anxiety and health related concerns is consistent with other work in this field (Domschke et al., 2010; Paulus & Stein, 2010). A noteworthy observation was that self-reported frequency of mindfulness practice (see 'Demographics') was positively correlated with interoceptive sensibility (r=.274, p=.001) suggesting people with a tendency to focus on bodily sensations engaged in mindfulness more often. Therefore, people exhibiting a greater tendency to notice bodily sensations in this study, may have demonstrated a healthier, more adaptive attentional style towards bodily sensations as opposed to a more anxious preoccupation with bodily sensations.

Table 2.

Pearson's coefficients for all independent variables

	Thirst	***State Anx.***	***Trait Anx***	***Pos Affect***	***Neg Affect***	***Intero.***	***CRT***	***Age***	***Sex***	***Health***
Hunger	.289**	-.30	-.089	-.031	-.071	.152	-.069	.000	.041	.050
Thirst		.165	.136	-.117	.140	.247**	.012	-.061	.064	-.154
State Anx.			.726**	-.197*	.454**	.038	-.045	-.318**	.119	-.430**
Trait Anx.				-.284**	.454**	.121	-.008	-.294**	.227**	-.325**
Pos. Affect					-.072	-.125	.057	.143	-.156	.204*
Neg. Affect						.083	-.023	-.205*	.230**	-.148
Intero.							-.216**	-.004	.093	-.072
CRT								.047	-.121	-.029
Age									-.253**	.155
Sex										-.040

Note. Variables in vertical order: Hunger, Thirst, State Anxiety, Trait anxiety, Positive Affect, Negative Affect, Interoceptive sensibility, Cognitive Reflection Task score (regular), Age, Sex (Male=1, Female=2), Current Health (* p<.05; **p<.01).

Analyses

R1. Hunger, thirst, and moral judgments of harm

R1 tested whether felt hunger or thirst biased moral judgments of harm. Hunger and thirst ratings were entered as predictor variables in four multiple linear regression models. Outcome-maximisation, harm-aversion and moral acceptability for congruent (acceptability_congruent) and incongruent trials (acceptability_incongruent) were the dependent variables. Contrary to our hypothesis, neither hunger (b=.072, p=.402) nor thirst

(*b*=.036, *p*=.675) predicted participant's harm aversion scores (R^2 *chng*=.010, *F* (2, 149) = 0.597, *p*=.552). Moreover, hunger (*b*=0.043) and thirst (*b*= -0.043) did not predict participant outcome-maximisation scores (R^2 *chng*=.003, *F* (2, 149) = 0.195, *p*=.823). Therefore, participants who were more hungry or thirsty were not more or less likely to accept/reject harm or maximise outcomes on the moral dilemma task than those who were less hungry or thirsty. Hunger (*b*=0.118, *p*=.170) and Thirst (*b*=-0.052, *p*=.544) also did not predict acceptability_incongruent scores (where harm results in a 'greater good') (R^2 *chng*=.013, *F* (2, 148) = 0.973, *p*=.380). Hunger did negatively predict acceptability _congruent scores (*b*=0.226, *p*=.008) in the model (R^2 *chng*=.049, *F* (2, 148) = 3.841, *p*=.024) but thirst did not (*b*=-0.116, *p*=.167). Therefore, hungrier participants were more likely to judge the moral acceptability of harmful actions as less 'wrong', but only for trials where harmful actions resulted in no greater-good overall. Finally, as hunger and thirst are related sensations often physiologically interlinked, a post hoc mediation analysis was conducted to assess whether hunger (*a*) influenced acceptability_congruent scores through experiences of thirst (*b*). A bootstrap confidence interval for the indirect effect (*ab*=-0.0421) included zero (-0.0059 to 0.0005) indicating that hunger did not influence moral acceptability ratings on congruent trials through related experiences of thirst. Despite the positive correlation between hunger and thirst ratings, how thirsty participants felt did not appear influential for moral acceptability ratings across all trials.

R2. Moderating role of affective state

In R2, we explored the moderating role of affective state in the relationship between hunger and moral judgments of harm. Against our predictions, no significant correlations were found between hunger, thirst and positive or negative emotional state, or between emotional state and moral judgments (see Table 2). Non-significant relationships between hunger/thirst and moral judgments were not probed for moderation effects of emotional state. A moderating role of positive affect (R^2 chng=.0118, *F* (1, 147) = 1.82, *p*=.1793) and negative affect (R^2 chng=.012, *F* (1, 147) = 1.853, *p*=.1755) was not found in the relationship found between hunger and acceptability_congruent ratings found in R1. A further mediation analysis was carried out to rule out the possibility of hunger influencing acceptability_congruent ratings through changes in emotional state. Bootstrap confidence intervals of the indirect effect of hunger through positive affect (-.0012 to .0012) and negative affect (-.0013 to .0008) on acceptability_congruent ratings were entirely below zero, ruling out any mediation effects. Together this indicates, the influence of hunger on moral

acceptability ratings of unprofitable harmful acts cannot be explained by hunger-associated changes in emotional state. A two-step hierarchical regression controlling for the effects of positive affect, negative affect and state anxiety also confirmed hunger significantly influenced acceptability_congruent ratings (b=.197, p=.016). Therefore, the influence of hunger on non-profitable judgments of harm was independent of affective experience.

R3. Moderating role of interoceptive sensibility

In R3 we proposed that greater interoceptive sensibility (tendency to focus on bodily sensations) could increase the availability of visceral sensations associated with hunger or thirst which could moderate the relationship between hunger/thirst and moral judgments. Contrary to R3, a moderation analysis yielded no interaction effect between hunger and interoception for outcome-maximisation tendencies (R^2 chng=<.001, F (1, 147) = 0.054, p=.817), harm-aversion tendencies (R^2 chng=.0179, F (1, 147) = 2.719; p=.101), acceptability_incongruent scores (R^2 chng=.0064, F (1, 147) = 0.964, p=.328) or acceptability_congruent scores (R^2 chng=.003, F (1, 147) = 0.455, p=.501). Similarly, no moderation effect of interoception was found between thirst and outcome-maximisation (R^2 chng=.000, F (1, 147) = 0.064, p=.799), harm-aversion tendencies (R^2 chng=.0026, F (1, 147) = 0.395, p=.531), acceptability_incongruent scores (R^2 chng=<.001, F (1, 147) = 0.002, p=.961), or acceptability_congruent scores (R^2 chng=<.001, F (1, 147) = 0.005, p=.979). Therefore, the influence of sensations of hunger or thirst on participant's moral acceptability and moral appropriateness judgments did not vary as a function of their tendency to focus on bodily sensations.

R4. Influence of sex, age, and anxiety on moral judgments

Age and sex were inputted as predictors in multiple linear regression models of all of the dependent variables (harm-aversion, outcome-maximisation tendencies, acceptability_congruent and acceptability_incongruent ratings). We found participants age (b=0.216, p=.009) and sex (b=0.220, p=.008) significantly predicted harm-aversion tendencies, with females and older participants showing greater harm-aversion tendencies irrespective of the consequences of harm (R^2=.071, F (2, 148) = 5.656, p=.004). Neither age nor sex predicted outcome-maximisation tendencies i.e. acceptance of harm in the interests of the 'greater good' (R^2=.003, F (2, 148) = 0.215, p=.807). Age and sex also did not predict acceptability_congruent (R^2=.022, F (2, 148) = 1.626, p=.200) or acceptability_incongruent ratings (R^2=.026, F (2, 148) = 1.959, p=.145). Therefore, age and sex did not influence how

morally acceptable people judged harmful actions to be, despite the differences in harm-aversion tendencies overall. In partial support of the role of anxiety in moral judgments, state anxiety negatively correlated with harm-aversion (r=-.177, p=.03), indicating people who were more anxious at the time of the experiment were less likely to reject causing harm in the moral dilemmas. However, a hierarchical regression model confirmed that state anxiety did not significantly predict harm-aversion scores when controlling for the effects of sex and age, (R^2 chng=.091, F (3, 147) = 4.911, b=-.150, p=.078). State anxiety was significantly negatively correlated with age (r=.318, p<.01), with younger people more likely to report both state and trait anxiety. Therefore, age appears to account for much of the variation in state anxiety that predicted harm rejection judgments towards the moral dilemmas.

Exploratory analyses

We tested the hypothesis that more accurate performance on the CRT task would positively predict more utilitarian response tendencies in line with prior research (Baron et al., 2015; Byrd & Conway, 2019).

EH1. CRT performance and moral judgments

Contrary to EH1, CRT performance showed significant positive correlation with harm-aversion (r=.235, p=.004) but not outcome-maximisation tendencies (r=.048, p=.562). As gender differences have been found for CRT performance (Ring et al., 2016) a multiple linear regression controlling for the effects of age and sex confirmed that CRT scores significantly predicted harm-aversion tendencies (R^2 chng=.135, F (3, 147) = 7.655, b=.255, p=.001). This finding was sustained when inputting alternative CRT scores representing the presence of 'intuitive' answers as opposed to correct answers. Therefore, participants who were more likely to provide 'intuitive' answers on the CRT were more likely to accept causing harm in moral dilemmas, irrespective of the outcomes.

Following EH1, we explored the predictive relationship between interoceptive sensibility and the performance on the CRT task. We further investigated the possibility of a mediation-effect of interoceptive sensibility through intuitive decision-making processes (captured by the CRT) on harm-aversion responses.

EH2. CRT performance, interoceptive sensibility and harm-aversion

Following identification of moderate correlation between CRT scores and interoceptive sensibility (r=-.216, p=.008), a linear regression model confirmed that higher interoceptive sensibility predicted more incorrect and intuitive responses on the CRT (R^2 chng=.046, F (1, 149) = 7.264, p=.008). A mediation analysis explored the presence of an indirect effect of interoceptive sensibility (*a*) on harm-aversion scores, through more 'intuitive' decision-making processes on the CRT (*b*). The direct effect of interoception on harm-aversion scores was not significant (t=-0.272, p=.786). However, a bootstrap confidence interval of the indirect effect (*ab*=-.0802) was entirely below zero (-0.1718 to -.0110) suggesting that people with a greater tendency to focus on internal sensations provided more intuitive responses on the CRT and were more likely to condone harmful actions. Therefore, interoceptive sensibility explains a significant amount of variance in 'intuitive' CRT responses which subsequently predicts participants acceptance of harmful actions.

Discussion

An unexpected and novel discovery in this study was that hunger-bias appeared uniquely influential for acceptability judgments of non-profitable harmful actions, whereas 'intuitive' decision-making tendencies exclusively predicted appropriateness judgments of harm. These independent effects suggest that a metaphorical 'gut instinct' and gut-related visceral experiences of hunger have distinct influences on harm-based moral cognition. We do have the capacity to be morally hypocritical; although we may judge an action to be morally appropriate, we can equally judge that act to be morally unacceptable (Tassy et al., 2013). Framing questions as '*is it appropriate to…?*' versus '*how morally acceptable do you find…?*' assumes different perspectives of the judge and inconsistencies have been found between these types of judgments previously (Pletti et al., 2017; Tassy et al., 2013). Choice judgments such as "*Would you do…in order to…?"* involves forming a judgment from an egocentric perspective and makes self-relevant consequences more salient (Sood & Forehand, 2005; Tassy et al., 2013). Choice judgments are akin to the moral appropriateness judgments in this study which encouraged people to adopt the perspective of the person carrying out the harmful action in the story (see Table 1). Whereas moral acceptability judgments provide a more abstract or allocentric perspective to evaluate a harmful act and create distance from the self and refer to the moral acceptability judgments in this study (Frith & De Vignemont, 2005). These two types of judgments may rely on distinct neural bases associated with

differing degrees of agency. For example, egocentric moral judgments, but not allocentric, have been associated with activation of the amygdala suggesting these judgments activate emotional processes associated with weighing up the consequences of our own actions for ourselves (Berthoz et al., 2006). Moreover, experiencing oneself as the cause of action has shown to activate areas of the anterior insula (Adolfi et al., 2017) whereas experiencing someone else as the cause of action is associated with activation of the inferior parietal cortex (Farrer & Frith, 2002). The importance of 'where we are' in relation to harm, has also come to light in virtual-reality studies that find discrepancies between hypothetical moral judgments people make and the harmful behaviours they perform when confronted with more realistic moral dilemmas (e.g. Francis et al., 2016).

In R1, we found less hungry people rated harm as more wrong in instances where harm did not result in any 'greater good' overall, although this predictive relationship was relatively weak. This suggests hunger may be uniquely influential for allocentric judgments about unprofitable harmful acts. In line with prior research, the physiological changes associated with hunger states could bias how severely we judge the acceptability of moral violations from an allocentric perspective (Vicario et al., 2018), but the exclusivity of this effect for non-profitable harmful actions is a novel finding for harm-based moral dilemmas. Arguably, acceptability judgments for the trials where harm did not result in a 'greater good' provides a judgment of the 'wrongness' of excessive harm, as there is no moral justification to judge harm that is without benefit as morally acceptable. However, clearly people did judge certain types of harm to be more acceptable than other types and appears to have been influenced by their level of hunger. This is discussed further below. Moral appropriateness judgments, however, may reflect more stable aversions people have to characteristically harmful actions (action-aversion) and witnessing the pain of others (outcome-aversion) (Miller et al., 2014), that are more impervious to temporary hunger states. Alternatively, the binary yes/no option, may simply prevent us from understanding the true strength of appropriateness judgments. Surprisingly, in R3 we found emotional state did not moderate the relationship between hunger and moral judgments. Interoceptive sensibility also did not moderate any relationships between hunger, thirst and moral judgments, contrary (R2). Null findings for R2 and R3 suggest that hunger 'acted alone' to influence non-profitable moral acceptability judgments of harm and cannot be explained by differences in people's tendency to focus on visceral sensations like hunger, or incidental emotional state (e.g. Valdesolo & Desteno, 2006)

Perhaps most surprising was that CRT performance (directly) and interoceptive sensibility (indirectly) predicted people's harm-aversion tendencies. This may suggest a discrete influence of intuitive decision-making processes and interoceptive sensibility for judgments of harm when adopting an egocentric viewpoint of the actor causing harm. This finding contradicts exploratory hypothesis EH1 and prior research showing a positive relationship between CRT performance and outcome-maximisation or 'utilitarian' tendencies (Baron et al., 2015; Byrd & Conway, 2019). Although a logical reflection measure has correlated with harm-aversion tendencies before, arithmetic reflection (assessed by the CRT) has not (Byrd & Conway, 2019). The fact that more arithmetically correct answers on the CRT predicted the rejection of harmful action when harm resulted in a 'greater good' and when it did not, challenges the view that a more arithmetic focus is responsible for moral judgments that prioritise the number of lives saved (Byrd & Conway, 2019; Patil et al., 2020). An association between CRT performance and harm-aversion is also counter-intuitive to dual-process perspectives (Greene et al., 2001) that propose the rejection of harmful actions is associated with a faster, more emotional decision-making pathway that we might expect to negatively correlate with intuitive responses on the CRT (e.g. Kahneman and Frederick, 2002). People with higher interoceptive sensibility were more likely to provide 'intuitive' answers on the CRT task, suggesting greater bodily awareness impeded successful performance on this task. Furthermore, in EH2 we found interoceptive sensibility indirectly predicted harm-aversion tendencies through its influence on CRT performance. Whereby, heightened interoceptive sensibility appeared to reduce people's ability to resolve counter-intuitive problems on the CRT, which subsequently increased the likelihood they would condone harmful actions on the moral dilemma task. There is some support for the notion that an awareness of somatic states could actually enhance our representations of ourselves in relation to our moral responsibilities (Immordino-Yang, 2011), but the findings here suggest a heightened focus on visceral sensations may somehow contribute to a weakening of our aversion to harmful actions.

In R4, we found age and sex were the strongest predictors of harm-aversion but not outcome-maximisation tendencies. As founds previously, older participants and female participants were most likely to reject causing harm (Armstrong et al., 2019; McNair et al., 2019), but these age and sex differences did not extend to moral acceptability ratings- a distinction that has not been clarified before. There was partial support for a role of state anxiety in predicting harm-aversion tendencies, with more anxious people more likely to

accept causing harm regardless of the outcomes. This is somewhat consistent with a previous finding (Kouchaki & Desai, 2015) and is potentially due to how anxiety influences how we process threatening information (see Garfinkel & Critchley 2016). However, the predictive value of state anxiety appeared to be mostly explained by variation in age, with younger people significantly more likely to report higher levels of state anxiety.

Hunger and moral acceptability ratings

We did find hungrier people were more likely to judge non-profitable harmful actions as more morally acceptable, although the magnitude of effect was relatively weak and should be interpreted with caution. An absence of a relationship between hunger and state anxiety suggests these appraisals were not based on hunger-induced arousal (e.g. Chan et al., 2007; Korbonits et al., 2004). Indeed, hunger may not always induce physiological arousal in a negative sense (e.g. Michalsen, 2010) and hunger and state anxiety were in fact slightly negatively correlated in this study. Psychophysiological arousal has shown to predict an aversion to harmful actions (Cushman et al., 2012). Therefore, if arousal cues were reduced in hungrier individuals it is possible this lessened the severity of their acceptability judgments and is consistent with the finding that hunger can actually reduce threat-tolerance and promote riskier decision-making in animals (Ghosh et al., 2016). As the majority of people reported lower levels of hunger, it is possible that our sample did not include enough 'very-hungry' participants to generate the hormonal and physiological responses associated with hunger-induced arousal. This subsequently reduces the probability of observing individual differences in state anxiety or negative affect associated with hunger that may have been influential.

As people who were less hungry reported to have eaten more recently, a 'fullness'-based explanation is perhaps more likely and is consistent with some prior research (Vicario et al., 2018). Nausea symptoms often correlate with post-eating gastric emptying (Halawi et al., 2017) and can be interpreted emotionally as disgust (Tracy et al., 2019) which can influence moral judgments (see Haidt, McCauley, & Rozin 1994). However, this is a novel finding for harm-based moral judgments (Horberg et al., 2009). Nevertheless, this explanation is consistent with research finding positive correlations between hunger and acceptance of moral violations (Vicario et al., 2018), and between disgust sensitivity and disapproval of moral violations (Horberg et al., 2009; Vicario & Rafal, 2017). Unfortunately, as we did not measure disgust or fullness these hypotheses remain speculative, although only

3% of the sample reported any nausea or gastrointestinal distress in a pre-study health questionnaire. Importantly, nausea associated with gastric dysrhythmias is not unique to visceral signaling processes associated with eating and can occur during hunger states and stomach emptiness (see Levine, 2005 for a review). However, why exactly hunger was only influential for moral acceptability judgments of harm that did not result in any 'greater good' overall is unclear and requires further investigation.

Interoceptive sensibility

Interoceptive sensibility (IS) did not moderate the relationship between hunger and acceptability judgments of non-profitable harm (R2). As people's tendency to focus on visceral sensations did not change the relationship between hunger and moral judgments, this implies that the psychophysiological processes proposed to underlie this relationship (e.g. Vicario et al., 2018) do not strengthen with higher levels of attention directed towards internal sensations. However, it is conceivable that people with higher levels of IS had lower thresholds for detecting sensations of hunger and thirst (Stevenson et al., 2015) and were more likely to overestimate 'true' homeostatic states of hunger/thirst. This was evident for thirst at least, as a moderate correlation between thirst and IS (r=.247, p<.01), indicated people with a greater sensitivity to bodily sensations were more aware of thirst-type visceral sensations. Similarly, hunger was positively correlated with IS (r=.152) but was non-significant. It could be argued, that the hunger and thirst ratings scales provided a measure of IS themselves, as they asked people to consciously assess and report subjective visceral states which will of course depend on the availability of this information. Although problematic levels of multicollinearity between hunger/thirst and IS were not identified in the regression analyses; if changes in IS were met with corresponding changes in hunger/thirst ratings this would reduce the likelihood of observing any moderation effects of IS in the relationship between hunger and moral judgments. Future work using a larger sample could generate more statistical power to uncover any small effect sizes of IS in the link between hunger and moral acceptability judgments not found here.

Role of emotional states and anxiety

Interestingly, neither anxiety nor emotional state correlated with hunger or thirst providing no support for any association between these constructs found previously (see MacCormack, 2016). Agreeing on the archetypical symptoms and psychophysiological experiences of hunger and thirst is challenging, due to variations in eating contexts (Ribeiro

et al., 2009) and the variation of visceral and emotional expressions of hunger and thirst people report (e.g. Michalsen, 2010). For example, a large proportion of people do not experience abdominal emptiness when hungry (Harris & Wardle, 1987) and some even report positive psychological experiences from food deprivation (Watkins & Serpell, 2016) which could partially explain why we did not find the anticipated relationships between hunger, thirst and emotional states in this study. Alternatively, the null finding for R3 is perhaps due to the low variation in hunger and thirst ratings in this sample.

State anxiety negatively predicted harm-aversion tendencies on the moral judgment task but fell from significance when controlling for age and sex. Trait anxiety did not correlate with harm-aversion tendencies which contradicts an earlier finding (Zhao et al., 2016), and is surprising considering that measures of dispositional threat-reactivity have predicted people's aversion to harmful actions (Cushman et al., 2012) with momentary anxiety inductions having less of an effect on moral judgments (Zhao et al., 2016). Anxiety can facilitate the processing of threatening information (Mathews, 1990) and increase anticipation of aversive outcomes (Paulus & Stein, 2010) and may explain the negative association found between state anxiety and harm-acceptance tendencies found here. More anxious people may have perceived the hypothetical recipients of harm to be more threatening or considered the option of not carrying out harm (i.e. doing nothing) to be the riskier option compared to less anxious people, but the link between anxiety, physiological arousal and moral judgments is likely much more complex. Moreover, as emotion and anxiety were measured before participants completed the moral dilemmas, we can only speculate that any incidental feelings of anxiety or emotion were experienced as unrelated to the task as opposed to a reaction to the potential consequences of their choices on the task (see Baumeister et al., 2012).

CRT, interoceptive sensibility and harm-aversion

While interoceptive accuracy (on a heartbeat detection task) has shown to influence CRT performance under certain conditions (Lugo et al., 2017), a relationship between interoceptive sensibility (IS) and CRT performance is novel. Empirical work surrounding IS and cognition is limited. There is some evidence to suggest IS can influence risk-taking behaviour (Salvato et al., 2019) but this does not appear related to impulsivity in decision-making (Herman, Critchley, & Duka, 2018). IS also indirectly predicted harm-aversion tendencies through its influence on CRT performance. Both IS (Paulus & Stein, 2010) and

egocentric moral judgments are associated with forms of self-referential processing (Sood & Forehand, 2005) which is one speculative explanation of the indirect association between IS and harm-aversion in this study. People scoring higher on IS may engage in self-referential processing to a greater extent, which possibly reduces their inclination to engage in computationally demanding decision-processes when faced with counterintuitive problems like the CRT. Therefore, these people may be more likely to rely on intuitive heuristics to form their answer (Kahneman & Frederick, 2002) which may also have consequences for moral decision-making. Consistent with this hypothesis, perhaps the most parsimonious account for why the CRT was a predictor of harm-aversion bias here is because it taps into our general tendency towards being 'cognitive misers' – preferring the processing option that requires least energy expenditure (Toplak et al., 2011; Tversky & Kahneman, 1974). More intuitive responses on the CRT task could suggest reduced engagement or deliberation of items across the whole experiment, including the moral dilemmas where harmful actions do not result in a greater good overall. For these 'congruent' dilemmas, weighing up the consequences of harmful action arguably requires slightly more scrutiny of the story content at times. If participants did not fully consider the specific content of the stories, they could have mistakenly condoned harmful actions due to misreading or overlooking story information, which would provide a negatively skewed measure of harm-aversion for these people. Therefore, rather than poorer performance on the CRT task representing stable differences in intuitive thinking styles, it is possibly more a reflection of intuitive 'preference' (Pennycook et al., 2016) based on the computational resource available or utilized at that moment (Toplak et al., 2011).

Sex and age effects

In line with prior research using traditional moral dilemma paradigms, older participants demonstrated greater harm-aversion preferences which has been linked to a greater propensity to experience negative emotions (McNair et al., 2019) and/or a reduced ability to overcome affective cues when making judgments (Hess et al., 2000). Older participants in this study reported lower negative affect and state/trait anxiety than younger participants, and no age-related differences were found for interoceptive sensibility. Therefore, incidental negative affect (unrelated to the task) or greater attentional focus towards affective cues in the body does not appear to underlie the finding here, but a more negative emotional response to the moral dilemma stimuli from older participants cannot be ruled out. Similarly, although some research has shown men demonstrate stronger utilitarian

preferences (Tinghög et al., 2016), the finding that females scored similarly to men on utilitarian preferences but higher on harm-aversion tendencies is in line with cumulative research findings in this field (Armstrong et al., 2019). Explanations for gender differences typically centre around differences in socialisation practices (Wood & Eagly, 2012) as well as evolutionary pressures and physiological differences (see Armstrong, Friesdorf, & Conway 2019 for a review) which may engender greater social and emotional responses to the prospect of harming others in women.

Limitations

Online research into hunger and thirst has the advantage of gathering data from people in their natural eating environments but does not guarantee variation in visceral experiences or the presence of real physiological changes associated with hunger and thirst. It is possible that low variability of hunger (3.24 ± 2.09) and thirst ratings (4.07 ± 1.96) prevented us from uncovering individual differences in the impact of visceral and emotional states on moral decision-making. In addition, relying on self-report measures cannot provide an objective understanding of the physiological conditions accompanying these subjective states, and some research has found intra-individual inconsistencies using visual analogue scales of appetite (e.g. Flint et al., 2000). One indication of reliability of our measure is that hunger significantly predicted hours since eating, providing the expected relationship between hunger states and reported ingestive behaviour. Although we can never know what hungry or thirsty 'feels' like to different people or guarantee a consistent impact of food-deprivation manipulations on visceral experiences (Michalsen, 2010; Stevenson et al., 2015), using fasting manipulations (Vicario et al., 2018) or measures of blood-glucose (Gailliot et al., 2007) would allow the objective investigation of the impact of homeostatic depletion on moral decision-making.

A second limitation was the measure of interoceptive sensibility used (PBCQ: (Miller, Murphy & Buss., 1981). Although popular in interoception research (e.g. Duschek et al., 2015; Ainley & Tsakiris, 2013; Garfinkel et al., 2015) the PBCQ provides a one-dimensional trait measure of perceptual awareness of bodily symptoms. An important distinction between body awareness attention styles (Mehling et al., 2018) is not captured by this measure and limits our understanding of interoceptive sensibility in this context. A more negative attentional style is associated with anxiety and somatization (Domschke et al., 2010; Ginzburg et al., 2014) whereas a more adaptive attentional focus on the body can enhance

self-regulatory processes associated with bodily sensations and is prevalent in mindfulness-style practices like body-scanning (Bornemann et al., 2014). As participants self-reported mindfulness practice positively correlated with interoceptive sensibility, it is possible that participants on the higher end of the interoceptive sensibility scale exhibited a 'healthier' attentional focus on bodily sensations which could explain the absence of any relationship between interoceptive sensibility and anxiety. Future work using a Multidimensional Assessment of Interoceptive Awareness (Mehling et al., 2018) would provide a more nuanced understanding of the attentional and emotional regulation styles of people with higher levels of interoceptive sensibility.

The CRT (Frederick, 2005) is a popular but controversial measure, inherently confounded with numeracy ability. It is possible the CRT provides an indication of people's tendency to think in less effortful ways as opposed to reflecting stable individual differences in thinking styles (Toplak et al., 2011). A recent study found CRT scores did not reflect thinking styles or intuitive ability that was distinct from a general intelligence measure (Blacksmith et al., 2019) whereas other research suggests the CRT is valid for measuring reflective but not intuitive thinking styles (Pennycook et al., 2016). Ambiguity about whether the CRT taps into stable psychological constructs or more temporary psychological processes, can make the interpretation of results difficult. Future replications could clarify whether the CRT's power in predicting harm-aversion judgments was due stable individual differences in intuitive or rational thinking styles using measures such as the Rational-Experiential Inventory (Pacini & Epstein, 1999). Finally, our sample was a moderate size and well-represented in terms of age but contained a disproportionate number of women. Considering small effect-sizes and several null findings in this study, a more substantial and representative sample would increase the power to uncover effects of visceral states and interoceptive processes on moral judgments if they do exist.

Conclusion

When making difficult moral decisions we may refer to a metaphorical 'gut instinct' to explain our choices; a feeling we locate in our stomach area that steers us one way or another. Hunger is one such sensation fundamentally linked with our gastrointestinal system that appears to play a role in allocentric judgments of harmful acts and other moral transgressions, potentially due to its link with disgust (Schnall et al., 2008; Vicario et al., 2018). We also associate 'gut feelings' with a felt sense of intuition. Intuition is easily linked

with interoceptive processes, when we cannot consciously access the homeostatic valuations happening between the brain and body that can bias our decision-making processes (Craig 2015; Damasio, Everitt, & Bishop, 1996). Here, we found intuitive thinking preferences on the CRT was associated with a tendency to pay attention to interoceptive sensations and a reduced aversion to harmful actions. Together, these findings suggest hunger-bias and intuitive thinking preferences may represent independent processes shaping different types of moral judgments. It is possible that the presence of 'intuitive' responses on the CRT, may instead represent an absence of deliberative thinking processes (e.g. Toplak, West, & Stanovich, 2011), and we speculated that increased monitoring of bodily sensations associated with body awareness, could interfere with more effortful thinking processes due to the demand on attentional resources. Further work using validated measures of intuitive thinking (e.g. Pacini & Epstein, 1999) could clarify this supposition. Interestingly, incidental emotion and anxiety states did not moderate any relationship between hunger, interoception, CRT performance and moral judgments. This suggests that emotional state at the time of making harm-based moral judgments did not provide any significant contribution to these effects, contrary to our hypotheses (Valdesolo & Desteno, 2006). Future work measuring people's emotional state before and after the task could clarify whether a change in emotional state is more predictive of moral judgments than incidental emotional state. The findings of this study have gone some way in clarifying the influence of incidental visceral states, emotion and interoceptive sensibility on moral judgments of harm. Interoception is significantly understudied in morality research which provides many more research opportunities to explore the complex relationships between interoceptive processes, emotion and moral decision-making.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

H.B. conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper.

Acknowledgments

We thank Dr Susie Martin (University of Bath) for support with Qualtrics setup and recruitment. We thank Dr Harry Farmer (University of Greenwich) for useful discussions. This studentship runs alongside UK EPSRC grant Virtual Realities: Immersive Documentary Encounters (ref: EP/P025595/1).

References

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., Manes, F., and Ibáñez, A. (2017). Convergence of interoception, emotion, and social cognition: A twofold fMRI meta-analysis and lesion approach. *Cortex* 88, 124–142. doi: 10.1016/j.cortex.2016.12.019

Ainley, V., and Tsakiris, M. (2013). Body Conscious? Interoceptive Awareness, Measured by Heartbeat Perception, Is Negatively Correlated with Self-Objectification. *PLoS One* 8(2),. doi: 10.1371/journal.pone.0055568

Anderson, S. W., Bechara, A., Damásio, H., Tranel, D., and Damásio, A. R. (2013). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Soc. Neurosci. Key Readings* 9780203496(11), 29–40. doi: 10.4324/9780203496190

Armstrong, J., Friesdorf, R., and Conway, P. (2019). Clarifying Gender Differences in Moral Dilemma Judgments: The Complementary Roles of Harm Aversion and Action Aversion. *Soc. Psychol. Personal. Sci.* 10(3), 353–363. doi: 10.1177/1948550618755873

Baron, J., Scott, S., Fincher, K., and Emlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *J. Appl. Res. Mem. Cogn.* 4(3), 265–284. doi: 10.1016/j.jarmac.2014.09.003

Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsw154

Barrett, L., Quigley, K. S., Bliss-Moreau, E., and Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional. *J Pers Soc Psychol.* 23(1), 1–7. doi: 10.1038/jid.2014.371

Baumeister, R., Vohs, K., Clore, G. L., and Bar-Anan, Y. (2012). Affect-As-Information. *Encycl. Soc. Psychol.* 121–144. doi: 10.4135/9781412956253.n8

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., and Dolan, R. J. (2006). Affective response to one's own moral violations. *Neuroimage* 31(2), 945–950. doi: 10.1016/j.neuroimage.2005.12.039

Blacksmith, N., Yang, Y., Behrend, T. S., and Ruark, G. A. (2019). Assessing the validity of inferences from scores on the cognitive reflection test. *J. Behav. Decis. Mak.* 32(5), 599–612. doi: 10.1002/bdm.2133

Bornemann, B., Herbert, B. M., Mehling, W. E., and Singer, T. (2014). Differential changes in self-reported aspects of interoceptive awareness through 3 months of contemplative training. *Front. Psychol.* 5(OCT),. doi: 10.3389/fpsyg.2014.01504

Bornemann, B., Herbert, B. M., Mehling, W. E., and Singer, T. (2015). Differential changes in self-reported aspects of interoceptive awareness through 3 months of contemplative training. *Front. Psychol.* 5, 1504. doi: 10.3389/fpsyg.2014.01504

Brown, H., Fraser, D. S., and Proulx, M. J. (2019). *Does felt hunger and interoceptive sensibility influence moral judgments of harm?* osf.io/37vsh

Byrd, N., and Conway, P. (2019). Not all who ponder count costs: Arithmetic reflection

predicts utilitarian tendencies, but logical reflection predicts both deontological and utilitarian tendencies. *Cognition* 192(June), 103995. doi: 10.1016/j.cognition.2019.06.007

Calder, A. J., Beaver, J. D., Davis, M. H., Van Ditzhuijzen, J., Keane, J., and Lawrence, A. D. (2007). Disgust sensitivity predicts the insula and pallidal response to pictures of disgusting foods. *Eur. J. Neurosci.* 25(11), 3422–3428. doi: 10.1111/j.1460-9568.2007.05604.x

Chan, J. L., Mietus, J. E., Raciti, P. M., Goldberger, A. L., and Mantzoros, C. S. (2007). Short-term fasting-induced autonomic activation and changes in catecholamine levels are not mediated by changes in leptin levels in healthy humans. *Clin. Endocrinol. (Oxf).* 66(1), 49–57. doi: 10.1111/j.1365-2265.2006.02684.x

Christensen, A. J., Wiebe, J. S., Edwards, D. L., Michels, J. D., and Lawton, W. J. (1996). Body consciousness, illness-related impairment, and patient adherence in hemodialysis. *J. Consult. Clin. Psychol.* 64(1), 147–152. doi: 10.1037/0022-006X.64.1.147

Clark, D. M., Salkovskis, P. M., Breitholtz, E., Westling, B. E., Öst, L. G., Koehler, K. A., Jeavons, A., and Gelder, M. (1997). Misinterpretation of body sensations in panic disorder. *J. Consult. Clin. Psychol.* 65(2), 203–213. doi: 10.1037/0022-006X.65.2.203

Conway, P., and Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *J. Pers. Soc. Psychol.* 104(2), 216–235. doi: 10.1037/a0031021

Craig, B. A. D. (2015). *How do you Feel?:An Interoceptive Moment with Your Neurobiological Self*. Princeton University Press.

Crawford, J. R., and Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* 43(3), 245–265. doi: 10.1348/0144665031752934

Critchley, H. D., and Garfinkel, S. N. (2018). The influence of physiological signals on cognition. *Curr. Opin. Behav. Sci.* 19, 13–18. doi: 10.1016/j.cobeha.2017.08.014

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personal. Soc. Psychol. Rev.* 17(3), 273–292. doi: 10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., and Mendes, W. B. (2012). Simulating murder: the aversion to harmful action. *Emotion* 12(1), 2–7. doi: 10.1037/a0025071

Damasio, A. R., Tranel, D., and Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* 41(2), 81–94.

Damasio, A. R. (1994). *Descartes' error: Emotion, Reason and the Human Brain.* Grosset/Putnam books.

Damasio, A. R. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]. *Philos. Trans. Biol. Sci.* 351(1346), 1413–1420. doi: 10.1098/rstb.1996.0125

Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci. U. S. A.* 108(17), 6889–6892. doi: 10.1073/pnas.1018033108

Domschke, K., Stevens, S., Pfleiderer, B., and Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clin. Psychol. Rev.* 30(1), 1–11. doi: 10.1016/j.cpr.2009.08.008

Dunn, B. D., Dalgleish, T., and Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neurosci. Biobehav. Rev.* 30(2), 239–271. doi: 10.1016/j.neubiorev.2005.07.001

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., and Dalgleish, T. (2010). Listening to Your Heart: How Interoception Shapes Emotion Experience and Intuitive Decision Making. *Psychol. Sci.* 21(12), 1835–1844. doi: 10.1177/0956797610389191

Duschek, S., Werner, N. S., Reyes del Paso, G. A., and Schandry, R. (2015). The Contributions of Interoceptive Awareness to Cognitive and Affective Facets of Body Experience. *J. Individ. Differ.* 36(2), 110–118. doi: 10.1027/1614-0001/a000165

Erceg, N., and Bubić, A. (2017). One test, five scoring procedures: different ways of approaching the cognitive reflection test. *J. Cogn. Psychol.* 29(3), 381–392. doi: 10.1080/20445911.2016.1278004

Farrer, C., and Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage* 15(3), 596–603. doi: 10.1006/nimg.2001.1009

Ferentzi, E., Drew, R., Tihanyi, B. T., and Köteles, F. (2018). Interoceptive accuracy and body awareness – Temporal and longitudinal associations in a non-clinical sample. *Physiol. Behav.* 184(August 2017), 100–107. doi: 10.1016/j.physbeh.2017.11.015

Flint, A., Raben, A., Blundell, J., and Astrup, A. (2000). Reproducibility, power and validity of visual analogue scales in assessment of appetite sensation in single test meal studies Obesity-associated arterial hypertension-pathophysiology and treatment View project Collaboration with NIHS on RNA-seq data View. *Artic. Int. J. Obes.* 24, 38–48. doi: 10.1038/sj.ijo.0801083

Foot, P. (2003). The Problem of Abortion and the Doctrine of the Double Effect. *Virtues and Vices* 5, 19–32. doi: 10.1093/0199252866.003.0002

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Anderson, G. G., and Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS One* 11(10),. doi: 10.1371/journal.pone.0164374

Frederick, S. (2005). Cognitive Reflection and Decision Making. *J. Econ. Perspect.* doi: 10.1257/089533005775196732

Frith, U., and De Vignemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome. *Conscious. Cogn.* 14(4), 719–738. doi: 10.1016/j.concog.2005.04.006

Gailliot, M. T. (2013). Hunger and Reduced Self-Control in the Laboratory and across the World: Reducing Hunger as a Self-Control Panacea. *Psychology* 04(01), 59–66. doi: 10.4236/psych.2013.41008

Gailliot, M. T., Baumeister, R. F., Dewall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., Brewer, L. E., and Schmeichel, B. J. (2007). Self-control relies on glucose as a limited

energy source: Willpower is more than a metaphor. *J. Pers. Soc. Psychol.* 92(2), 325–336. doi: 10.1037/0022-3514.92.2.325

Garfinkel, S. N., and Critchley, H. D. (2013). Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: "Anterior insular cortex mediates bodily sensibility and social anxiety" by Terasawa et al.,(2012). *Soc. Cogn. Affect. Neurosci.* 8(3), 231–234. doi: 10.1093/scan/nss140

Garfinkel, S. N., and Critchley, H. D. (2016). Threat and the Body: How the Heart Supports Fear Processing. *Trends Cogn. Sci.* 20(1), 34–46. doi: 10.1016/j.tics.2015.10.005

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., and Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biol. Psychol.* 104, 65–74. doi: 10.1016/j.biopsycho.2014.11.004

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R. and Hüttter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion* 18(7), 989–1008.

Ghosh, D. D., Sanders, T., Hong, S., McCurdy, L. Y., Chase, D. L., Cohen, N., Koelle, M. R., and Nitabach, M. N. (2016). Neural Architecture of Hunger-Dependent Multisensory Decision Making in C. elegans. *Neuron* 92(5), 1049–1062. doi: 10.1016/j.neuron.2016.10.030

Ginzburg, K., Tsur, N., Barak-Nahum, A., and Defrin, R. (2014). Body awareness: Differentiating between sensitivity to and monitoring of bodily signals. *J. Behav. Med.* 37(3), 564–575. doi: 10.1007/s10865-013-9514-9

Gomez, B. T., Hansford, T. G., and Krause, G. A. (2007). The pepublicans should pray for rain: Weather, turnout, and voting in U.S. presidential elections. *J. Polit.* 69(3), 649–663. doi: 10.1111/j.1468-2508.2007.00565.x

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44(2), 389–400. doi: 10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An

fMRI Investigation of Emotional Engagement in Moral Judgment. *Science (80-.).* 293(5537), 2105 LP – 2108. http://science.sciencemag.org/content/293/5537/2105.abstract

Gu, J., Zhong, C. B., and Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *J. Exp. Psychol. Gen.* 142(2), 307–312. doi: 10.1037/a0029549

Häfner, M. (2013). When body and mind are talking: Interoception moderates embodied cognition. *Exp. Psychol.* 60(4), 255–259. doi: 10.1027/1618-3169/a000194

Haidt, J., McCauley, C., and Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Pers. Individ. Dif.* 16(5), 701–713. doi: 10.1016/0191-8869(94)90212-7

Halawi, H., Camilleri, M., Acosta, A., Vazquez-Roque, M., Oduyebo, I., Burton, D., Busciglio, I., and Zinsmeister, A. R. (2017). Relationship of gastric emptying or accommodation with satiation, satiety, and postprandial symptoms in health. *Am. J. Physiol. - Gastrointest. Liver Physiol.* 313(5), G442–G447. doi: 10.1152/ajpgi.00190.2017

Harris, A., and Wardle, J. (1987). The feeling of hunger. *Br. J. Clin. Psychol.* 26(2), 153–154. doi: 10.1111/j.2044-8260.1987.tb00745.x

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd ed.). The Guildford Press: New York.

Hellström, P. M., and Näslund, E. (2001). Interactions between gastric emptying and satiety, with special reference to glucagon-like peptide-1. *Physiol. Behav.* 74(4–5), 735–741. doi: 10.1016/S0031-9384(01)00618-7

Herbert, B. M., Herbert, C., Pollatos, O., Weimer, K., Enck, P., Sauer, H., and Zipfel, S. (2012). Effects of short-term food deprivation on interoceptive awareness, feelings and autonomic cardiac activity. *Biol. Psychol.* 89(1), 71–79. doi: 10.1016/j.biopsycho.2011.09.004

Herman, A. M., Critchley, H. D., and Duka, T. (2018). Risk-taking and impulsivity: The role

of mood states and interoception. *Front. Psychol.* 9(AUG), 1–11. doi: 10.3389/fpsyg.2018.01625

Herman, C. P., Polivy, J., Lank, C. N., and Heatherton, T. F. (1987). *Anxiety , Hunger , and Eating Behavior*. 96(3), 264–269.

Hess, T. M., Waters, S. J., and Bolstad, C. A. (2000). *Motivational and Cognitive Influences on Affective Priming in Adulthood*. 55(4), 193–204.

Horberg, E. J., Oveis, C., Keltner, D., and Cohen, A. B. (2009). Disgust and the Moralization of Purity. *J. Pers. Soc. Psychol.* 97(6), 963–976. doi: 10.1037/a0017423

Immordino-Yang, M. H. (2011). Me, my "self" and you: Neuropsychological relations between social emotion, self-awareness, and morality. *Emot. Rev.* 3(3), 313–315. doi: 10.1177/1754073911402391

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., and Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychol. Rev.* 125(2), 131–164. doi: 10.1037/rev0000093

Kahneman, D., and Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). Cambridge University Press.

Korbonits, M., Goldstone, A. P., Gueorguiev, M., and Grossman, A. B. (2004). Ghrelin - A hormone with multiple functions. *Front. Neuroendocrinol.* 25(1), 27–68. doi: 10.1016/j.yfrne.2004.03.002

Kouchaki, M., and Desai, S. D. (2015). Anxious, threatened, and also unethical: How anxiety makes individuals feel threatened and commit unethical acts. *J. Appl. Psychol.* 100(2), 360–375. doi: 10.1037/a0037796

Levine, M. E. (2005). Sickness and satiety: Physiological mechanisms underlying perceptions of nausea and stomach fullness. *Curr. Gastroenterol. Rep.* 7(4), 280–288. doi: 10.1007/s11894-005-0020-2

Linder, Jeffrey A., Jason N. Doctor, Mark W. Friedberg, Harry Reyes Nieva, Caroline Birks,

Daniella Meeker, and C. R. F. (2014). Time of day and the decision to prescribe antibiotics. *JAMA Intern. Med.* 174(12), 2029–2031. doi: 10.1001/jamainternmed.2014.5225.Author

Lugo, R. G., Sütterlin, S., Knox, B. J., Jøsok, Ø., Helkala, K., and Lande, N. M. (2017). The moderating influence of self-efficacy on interoceptive ability and counterintuitive decision making in officer cadets. *J. Mil. Stud.* 7(1), 44–52. doi: 10.1515/jms-2016-0005

MacCormack, J. (2016). Feeling "Hangry"? When hunger is conceptualized as an emotion. *Emotion* 19(2), 301.

Macht, M., and Simons, G. (2000). Emotions and eating in everyday life. *Appetite* 35(1), 65–71. doi: 10.1006/appe.2000.0325

Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behav. Res. Ther.* 28(6), 455–468. doi: 10.1016/0005-7967(90)90132-3

McGaugh, J. L. (2015). Consolidating Memories. *Annu. Rev. Psychol.* 66(1), 1–24. doi: 10.1146/annurev-psych-010814-014954

Mckiernan, F., Houchins, J. A., & Mattes, R. D. (2008). Relationships between human thirst, hunger, drinking, and feeding. *Physiology and Behavior*, *94*(5), 700–708.

McNair, S., Okan, Y., Hadjichristidis, C., and de Bruin, W. B. (2019). Age differences in moral judgment: Older adults are more deontological than younger adults. *J. Behav. Decis. Mak.* 32(1), 47–60. doi: 10.1002/bdm.2086

Mehling, W. E., Acree, M., Stewart, A., Silas, J., and Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PLoS One* 13(12), 1–12. doi: 10.1371/journal.pone.0208034

Michalsen, A. (2010). Prolonged fasting as a method of mood enhancement in chronic pain syndromes: A review of clinical evidence and mechanisms. *Curr. Pain Headache Rep.* 14(2), 80–87. doi: 10.1007/s11916-010-0104-z

Miller, L.C., Murphy, R. and Buss, A. H. (1981). Consciousness of body: Private and public. *J. Pers. Soc. Psychol.* 41(2), 397.

Miller, R. M., Hannikainen, I. A., and Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion* 14(3), 573—587. doi: 10.1037/a0035361

Moretto, G., Làdavas, E., Mattioli, F., and Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J. Cogn. Neurosci.* 22(8), 1888–1899. doi: 10.1162/jocn.2009.21367

Pacini, R., and Epstein, S. (1999). The Relation of Rational and Experiential Information Processing Styles to Personality, Basic Beliefs, and the Ratio-Bias Phenomenon. *J. Personal. Soc. Psychol.* 76(6), 972–987. http://10.0.4.13/0022-3514.76.6.972

Palomba, D., & Stegagno, L. (1995). Dissociation between actual and expected cardiac changes: Interoception and emotional experience. In *In D. Vaitl & R. Schandry (Eds.), From the heart to the brain: The psychophysiology of circulation – brain interaction* (pp. 283–298). Peter Lang Publishing.

Park, G., Kappes, A., Rho, Y., and Van Bavel, J. J. (2016). At the heart of morality lies neuro-visceral integration: Lower cardiac vagal tone predicts utilitarian moral judgment. *Soc. Cogn. Affect. Neurosci.* 11(10), 1588–1596. doi: 10.1093/scan/nsw077

Parton, D. M., and McGinley, J. J. (2019). At the heart of harm: Cardiac substrates of action-based aversion to harm. *Pers. Individ. Dif.* 151(July 2018), 109497. doi: 10.1016/j.paid.2019.07.007

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., and Cushman, F. (2020). Reasoning Supports Utilitarian Resolutions to Moral Dilemmas Across Diverse Measures. *J. Pers. Soc. Psychol.* 1–42. doi: 10.1037/pspp0000281

Paulus, M. P., and Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Struct. Funct.* 214(5–6), 451–463. doi: 10.1007/s00429-010-0258-9

Pennycook, G., Cheyne, J. A., Koehler, D. J., and Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behav. Res. Methods* 48(1), 341–348. doi: 10.3758/s13428-015-0576-1

Pletti, C., Lotto, L., Buodo, G., and Sarlo, M. (2017). It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. *Br. J. Psychol.* 108(2), 351–368. doi: 10.1111/bjop.12205

Ramsay, D., and Booth, D. (2012). *Thirst: physiological and psychological aspects.* (Eds.). Springer Science and Business Media.

Ribeiro, A. C., Lesauter, J., Dupré, C., and Pfaff, D. W. (2009). Relationship of arousal to circadian anticipatory behavior: Ventromedial hypothalamus: One node in a hunger-arousal network. *Eur. J. Neurosci.* 30(9), 1730–1738. doi: 10.1111/j.1460-9568.2009.06969.x

Ring, P., Neyse, L., David-Barett, T., and Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Front. Psychol.* 7(NOV), 1–7. doi: 10.3389/fpsyg.2016.01680

Rochette, L. M., and Patterson, S. M. (2005). Hydration status and cardiovascular function: Effects of hydration enhancement on cardiovascular function at rest and during psychological stress. *Int. J. Psychophysiol.* 56(1), 81–91. doi: 10.1016/j.ijpsycho.2004.10.003

Russell, J. A., and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Personal. Soc. Psychol.* 76(5), 805.

Salvato, G., De Maio, G., and Bottini, G. (2019). Interoceptive sensibility tunes risk-taking behaviour when body-related stimuli come into play. *Sci. Rep.* doi: 10.1038/s41598-019-39061-0

Sarker, M. R., Franks, S., and Caffrey, J. (2013). Direction of post-prandial ghrelin response associated with cortisol response, perceived stress and anxiety, and self-reported coping and hunger in obese women. *Behav. Brain Res.* 257, 197–200. doi: 10.1016/j.bbr.2013.09.046

Sarlo, M., Lotto, L., Rumiati, R., and Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiol. Behav.* 130, 127–134. doi: 10.1016/j.physbeh.2014.04.002

Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., and Sullivan, D. (2012). Concentrations on Human Decision-Making Performance. *Environ. Health Perspect.* 120(12), 1671–1678.

Schnall, S., Haidt, J., Clore, G. L., and Jordan, A. H. (2008). Disgust as embodied moral judgment. *Pers. Soc. Psychol. Bull.* 34(8), 1096–1109. doi: 10.1177/0146167208317771

Schwabe, L., Szinnai, G., Keller, U., and Schachinger, H. (2007). Dehydration does not influence cardiovascular reactivity to behavioural stress in young healthy humans. *Clin. Physiol. Funct. Imaging* 27(5), 291–297. doi: 10.1111/j.1475-097X.2007.00750.x

Sood, S., and Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organ. Behav. Hum. Decis. Process.* 98(2), 144–154. doi: 10.1016/j.obhdp.2005.05.005

Spielberger, C. D., and Gorsuch, R. L. (1983). *The state trait anxiety inventory for adults: sampler set manual, test, scoring key*. Redwood City Calif: Mind Garden.

Stevenson, R. J., Mahmut, M., and Rooney, K. (2015). Individual differences in the interoceptive states of hunger, fullness and thirst. *Appetite* 95, 44–57. doi: 10.1016/j.appet.2015.06.008

Sze, J. A., Gyurak, A., Yuan, J. W., and Levenson, R. W. (2010). Coherence Between Emotional Experience and Physiology: Does Body Awareness Training Have an Impact? *Emotion* 10(6), 803–814. doi: 10.1037/a0020146

Szekely, R. D., and Miu, A. C. (2015). Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cogn. Emot.* 29(1), 64–75. doi: 10.1080/02699931.2014.895300

Tassy, S., Oullier, O., Mancini, J., and Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Front. Psychol.* 4, 250. doi: 10.3389/fpsyg.2013.00250

Thomson, J. J. (1985). The trolley problem. *Yale Law J.* 94(6), 1395–1415.

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., and Västfjäll, D. (2016). Intuition and moral decision-making-the effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLoS One* 11(10), 1–19. doi: 10.1371/journal.pone.0164012

Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Mem. Cogn.* 39(7), 1275–1289. doi: 10.3758/s13421-011-0104-1

Tracy, J. L., Steckler, C. M., and Heltzel, G. (2019). The physiological basis of psychological disgust and moral judgments. *J. Pers. Soc. Psychol.* 116(1), 15–32. doi: 10.1037/pspa0000141

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (80-.).* 185(4157), 1124–1131. doi: 10.4324/9781912282562

Valdesolo, P., and Desteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychol. Sci.* 17(6), 476–477. doi: 10.1111/j.1467-9280.2006.01731.x

Vicario, C. M., Kuran, K. A., Rogers, R., and Rafal, R. D. (2018). The effect of hunger and satiety in the judgment of ethical violations. *Brain Cogn.* 125(February), 32–36. doi: 10.1016/j.bandc.2018.05.003

Vicario, C. M., and Rafal, R. D. (2017). Relationship between body mass index and moral disapproval rating for ethical violations. *Pers. Individ. Dif.* 104, 8–11. doi: 10.1016/j.paid.2016.07.027

Watkins, E., and Serpell, L. (2016). The Psychological Effects of Short-Term Fasting in Healthy Women. *Front. Nutr.* 3(August), 1–6. doi: 10.3389/fnut.2016.00027

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *J. Pers. Soc. Psychol.* 54(6), 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Weinshall-Margel, K., and Shapard, J. (2011). Overlooked factors in the analysis of parole decisions. *Proc. Natl. Acad. Sci. U. S. A.* 108(42), 2011. doi: 10.1073/pnas.1110910108

Werner, N. S., Duschek, S., Mattern, M., and Schandry, R. (2009). The relationship between pain perception and interoception. *J. Psychophysiol.* 23(1), 35–42. doi: 10.1027/0269-8803.23.1.35

Wheatley, T., and Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* 16(10), 780–784. doi: 10.1111/j.1467-9280.2005.01614.x

Wood, W., and Eagly, A. H. (2012). Biosocial Construction of Sex Differences and Similarities in Behavior. In *Advances in Experimental Social Psychology* (1st ed., Vol. 46). Elsevier Inc. doi: 10.1016/B978-0-12-394281-4.00002-7

Yam, K. C., Reynolds, S. J., and Hirsh, J. B. (2014). The hungry thief: Physiological deprivation and its effects on unethical behavior. *Organ. Behav. Hum. Decis. Process.* 125(2), 123–133. doi: 10.1016/j.obhdp.2014.07.002

Zaki, J., Davis, J. I., and Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *Neuroimage* 62(1), 493–499. doi: 10.1016/j.neuroimage.2012.05.012

Zhang, L., Kong, M., and Li, Z. (2017). Emotion regulation difficulties and moral judgment in different domains: The mediation of emotional valence and arousal. *Pers. Individ. Dif.* 109, 56–60. doi: 10.1016/j.paid.2016.12.049

Zhang, L., Li, Z., Wu, X., and Zhang, Z. (2017). Why people with more emotion regulation difficulties made a more deontological judgment: The role of deontological inclinations. *Front. Psychol.* doi: 10.3389/fpsyg.2017.02095

Zhao, J., Harris, M., and Vigo, R. (2016). Anxiety and moral judgment: The shared deontological tendency of the behavioral inhibition system and the unique utilitarian tendency of trait anxiety. *Pers. Individ. Dif.* 95, 29–33. doi: 10.1016/j.paid.2016.02.0

Supplementary material

Data processing and quality

Assumption checks of the data are presented in the results section. All data was appropriately distributed, except where mentioned and was corrected for one of the dependent variables. We observed expected relationships between trait/state anxiety and positive/negative effect suggesting the reliability of these measures. We also observed comparable internal consistency scores for the questionnaire measures as found in previous studies. Hunger and thirst measures were corroborated by hours since eating and drinking, suggesting this data was reliable. An attention check was not used in this study, and therefore, we cannot rule out the possibility of poor attention influencing performance on the moral dilemma task.

Moral Dilemma Items

Incongruent Time Machine

You find a time machine and travel back to the year 1920. While checking into a hotel, you meet a young Austrian artist and veteran of the First World War. You realize this is Adolf Hitler before his rise to power in Nazi Germany.

He is staying in the hotel room next to yours and the doors are not locked. It would be easy to simply smother him with a pillow in his sleep and disappear, stopping the Second World War and the Nazi party before they even start. However, he has not committed any crimes yet and it seems wrong to hurt an innocent person.

Is it appropriate for you to kill an innocent young Hitler in order to prevent the Second World War?

Congruent Time Machine

You find a time machine and travel back to the year 1920. While checking into a hotel, you meet a young petty criminal. You realize this is George Brackman, a man who later on abducted a child and held her for a week until her family paid him some ransom money.

He is staying in the hotel room next to yours and the doors are not locked. It would be easy to simply smother him with a pillow in his sleep and disappear, stopping the abduction and ransom demands before they even start. However, he has not committed any crimes yet and it seems wrong to hurt an innocent person.

Is it appropriate for you to kill George Brackman in order to prevent him from taking a child hostage?

Incongruent Car Accident
You are driving through a busy city street when all of a sudden a young mother carrying a child trips and falls into the path of your vehicle. You are going too fast to brake in time; your only hope is to swerve out of the way.

Unfortunately, the only place you can swerve is currently occupied by a little old lady. If you swerve to avoid the young mother and baby, you will seriously injure or kill the old lady.

Is it appropriate to swerve and hit the old lady in order to avoid the young mother and child?

Congruent Car Accident
You are driving through a busy city street when all of a sudden a young mother carrying a child trips and falls into the path of your vehicle. You are going too fast to brake in time; your only hope is to swerve out of the way.

Unfortunately, the only place you can swerve is currently occupied by a group of children on their way to elementary school. If you swerve to avoid the young mother and baby, you will seriously injure or kill several of them.

Is it appropriate to swerve and hit the schoolchildren in order to avoid the young mother and child?

Incongruent Hard Times
You are the head of a poor household in a developing country. Your crops have failed for the second year in a row, and it appears that you have no way to feed your family. Your sons, ages eight and ten, are too young to go off to the city where there are jobs, but your daughter could fare better.

You know a man from your village who lives in the city and who makes sexually explicit films featuring girls such as your daughter. In front of your daughter, he tells you that in one year of working in his studio your daughter could earn enough money to keep your family fed for several growing seasons.

Is it appropriate for you to employ your daughter in the pornography industry in order to feed your family?

Congruent Hard Times
You are the head of a poor household in a developing country. Your sons, ages eight and ten, work on the family farm everyday in order to teach them the value of hard work and providing for others. Your daughter, however, cannot work on the farm due to the heavy lifting required. You worry that she will not learn the value of money if she does not work.

You know a man from your village who lives in the city and who makes sexually explicit films featuring girls such as your daughter. In front of your daughter, he tells you that in one year of working in his studio your daughter could earn enough money to keep your family fed for several growing seasons.

Is it appropriate for you to employ your daughter in the pornography industry in order to teach her the value of money?

Incongruent Crying Baby

It is war time. Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables.

A baby with no parents begins to cry loudly. You cover her mouth to block the sound. If you remove your hand from the baby's mouth her crying will summon the attention of the soldiers who will kill you and the others hiding out in the cellar. To save yourself and the others you must smother the child to death.

Is it appropriate for you to smother the child in order to save yourself and the other townspeople from being killed?

Congruent Crying Baby

It is war time. Enemy soldiers have taken over your village. They have orders to capture all remaining civilians to make them work quarrying stone in a mine. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables.

A baby with no parents begins to cry loudly. You cover her mouth to block the sound. If you remove your hand from her mouth the crying will summon the attention of the soldiers who will capture you and the others hiding out in the cellar. To save yourself and the others from laboring in the mine you must smother the child to death.

Is it appropriate for you to smother the child in order to save yourself and the other townspeople from being captured?

Incongruent Relationship

You are divorced now and your former spouse lives far away. You don't expect to see your former spouse for a very long time. However, you have recently started dating a new person that is positively wonderful. Your new partner is head over heels in love for the first time and you feel equally strong about your partner. You want to do everything in your power to keep the relationship progressing smoothly to what you believe will be a second marriage.

Unfortunately, your new partner has told you many times that dating someone who is divorced is totally unacceptable. It is expressly forbidden in your partner's religion. Your partner doesn't know about your ex-spouse. You consider simply pretending that you have never been divorced so that your current relationship can continue.

Is it appropriate for you to leave your new partner in the dark about your previous relationship in order to keep her/him happy and the relationship alive?

Congruent Relationship

You are divorced now and your former spouse lives far away. You don't expect to see your former spouse for a very long time. However, you have recently started dating a new person that is positively wonderful. Your new partner is head over heels in love for the first time and you feel equally strong about your partner. You want to do everything in your power to keep the relationship progressing smoothly to what you believe will be a second marriage.

Unfortunately, your ex-spouse gave you a very serious and incurable Sexually Transmitted Disease (STD). Your partner does not know about it. You consider simply pretending that you don't know about the STD so that your current relationship can continue.

Is it appropriate for you to leave your new partner in the dark about your STD in order to keep her/him happy and the relationship alive?

Incongruent Abortion
You are a surgeon. A young woman you know becomes pregnant, but her body reacts in an unusual fashion. She develops a severe case of preeclampsia, a dangerous syndrome that leads to rapid increases in blood pressure. The only treatment is to deliver the baby.

Unless the baby is delivered soon, the mother will die. However, the baby is too young to survive on its own. If it is delivered, it will die. So, although it is very difficult for her, the mother asks you to abort the baby.

Is it appropriate for you to perform an abortion in order to save the mother's life?

Congruent Abortion
You are a surgeon. A young woman you know becomes pregnant, but she is not yet ready for children. She has not finished high school, has no income, and was abandoned by the father.

If she has the baby now, she will be stuck as a single mother on welfare for the rest of her life. This will make things very hard on her and the baby. She thinks that it would be smarter to wait and have children later. So, although it is very difficult for her, she asks you to abort the baby.

Is it appropriate for you to perform an abortion in order to let the mother live a better life?

Incongruent Torture
You are a police officer, and have recently caught a criminal you have been hunting for some time. He is allegedly responsible for rigging a series of explosive devices: some that have already gone off and some that have yet to detonate.

He places explosives outside city cafes and sets them to go off at a time when people are drinking coffee on the patios. In this manner, he has injured many people and might injure many more.

Now that the criminal is in custody, you want to know where the unexploded bombs are so you can defuse them. He refuses to talk, so you decide to use "aggressive interrogation techniques" like holding his head under water and beating him.

Is it appropriate for you to use "aggressive interrogation techniques" in order to find and defuse the unexploded bombs?

Congruent Torture
You are a police officer, and have recently caught a criminal you have been hunting for some time. He is allegedly responsible for rigging a series of explosive devices: some that have already gone off and some that have yet to detonate.

He places explosives outside city cafes and sets them to go off at a time when no one is around. His explosives are inside paint cans so that they spray nearby objects with paint. In this manner, he has sprayed many cafes with paint and might spray many more.

Now that the criminal is in custody, you want to know where the unexploded bombs are so you can defuse them. He refuses to talk, so you decide to use "aggressive interrogation techniques" like holding his head under water and beating him.

Is it appropriate for you to use "aggressive interrogation techniques" in order to find and defuse the unexploded bombs?

Incongruent Vaccine Policy

You are a doctor in a health clinic overrun by patients with a serious disease. You just received a shipment of drugs that can cure the disease but the drugs have their own severe side-effects.

If you administer the drugs to your patients, a small number will die from the side effects, but most will live. If you do not, most will die from the disease.

Is it appropriate for you to administer the drug to your patients?

Congruent Vaccine Policy

You are a doctor in a health clinic overrun by patients with the latest flu virus. You just received a shipment of drugs that can cure the flu but the drugs have their own severe side-effects.

If you administer the drugs to your patients, a small number will die from the side effects, but most will live. If you do not, most will continue to suffer from the effects of the flu virus for some time.

Is it appropriate for you to administer the drug to your patients?

Incongruent Animal Research

You have been hired by a pharmaceutical company to conduct research on their products. Since products must be fit for human use, they are first tried out on animals. Your job is to find out the effects various chemicals have on rats, pigeons, rabbits, and monkeys. Most chemicals have only minor effects on the animals, but some cause them discomfort or even permanent damage.

The chemicals you are researching are slated to form part of a new AIDS drug cocktail that will give new hope to millions of AIDS victims around the world. You anticipate saving many lives with the chemicals.

Is it appropriate to test these chemicals on animals?

Congruent Animal Research

You have been hired by a pharmaceutical company to conduct research on their products. Since products must be fit for human use, they are first tried out on animals. Your job is to

find out the effects various chemicals have on rats, pigeons, rabbits, and monkeys. Most chemicals have only minor effects on the animals, but some cause them discomfort or even permanent damage.

The chemicals you are researching are slated to form part of a new acne facial cleanser that will give new hope to people with pimples and greasy skin. You anticipate making many people feel better about their appearance with the chemicals.

Is it appropriate to test these chemicals on animals?

Incongruent Border Crossing

You are a soldier guarding a border checkpoint between your nation and one troubled by insurgent violence. You notice a young man in a cheap car approaching the checkpoint with a determined look on his face.

You suspect he means to bomb the checkpoint, killing all the soldiers inside. He is rapidly approaching your station.

Is it appropriate for you to shoot and kill the approaching man?

Congruent Border Crossing

You are a soldier guarding a border checkpoint between your nation and one troubled by insurgent violence. You notice a young man in a cheap car approaching the checkpoint with a determined look on his face.

You suspect he means to cross the border in order to work illegally inside your country. He is rapidly approaching your station.

Is it appropriate for you to shoot and kill the approaching man?

Study 1 Postface

Paper:

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). Hunger Bias or Gut Instinct ? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making. *Frontiers in Psychology: Perception Science*, *11*, 1–16. https://doi.org/10.3389/fpsyg.2020.02261

The key finding from this paper was the different associations between hunger, interoceptive sensibility and egocentric and allocentric judgments of harm. A lack of hunger enhanced the disapproval of allocentric judgments of unprofitable harmful acts but did not influence egocentric choice judgments. Whereas, interoceptive sensibility (indirectly) and a tendency towards intuitive decision-making (directly) predicted egocentric harm-aversive moral judgments. These findings were useful to reflect on in the context of Study 2 and 3, as they suggested the relationship between interoception, and moral judgment may be more evident for egocentric than allocentric judgments. Furthermore, the findings suggest that momentary hunger induced by pre-task protocols of e.g. not drinking/eating for several hours for interoceptive tasks may be less of a confounding factor for egocentric moral judgments. Although, we have since considered that measures of hunger and thirst may relate to constructs of IS as it requires people to consciously report interoceptive sensations which may have confounded our ability to find an interaction between hunger and IS. Surprisingly, thirst had no effect on any of the dependent variables, despite its relationship to hunger which suggests that interoceptive sensations of thirst did not as easily translate to emotional states in the way that gastric sensations associated with satiety/hunger do for disgust sensations and moral judgments (Tracy et al., 2019).

A key limitation of this study was that most of the sample reported being relatively less hungry/thirsty, which is unsurprising as people are probably less likely to participate in an online study when their homeostatic needs are significantly depleted. A lab-based between-subjects design using short-term fasting manipulations and potentially other measures of interoception would greatly improve our understanding of hunger, thirst, and

moral judgments of harm, although could potentially give way to demand characteristics. Nevertheless, this study revealed that even small differences in both the presence of interoceptive sensations and our habitual patterns of attending to bodily sensations can shape momentary moral judgments relating to harmful acts. However, overall, it appeared that hunger, thirst and interoceptive sensations only showed small effects on moral decision-making during harm-based moral dilemmas. Therefore, individual differences in momentary interoceptive state and hunger and thirst sensations may not be that influential for harm-based moral decision-making on moral dilemma tasks. This work provided useful 'food for thought' for the following studies and was rated as having wide audience-appeal during peer-review.

Study 2 Preface

Paper:

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). *Do individual differences in interoception influence the relationship between physiological arousal and moral judgments of harm? [Unpublished thesis]*. University of Bath.

Pre-registration:

Brown, H., Proulx, M. J., & Fraser, D. S. (2019). *Does interoception predict moral judgments of harm?* Open Science Framework. https://osf.io/m6f5j

Draft paper to be submitted to:

Cognitive, Affective and Behavioural Neuroscience

Study 2 aimed to assess whether individual differences in interoception influenced moral judgments of harm, but specifically whether interoception moderated the relationship between physiological arousal and moral judgments of harm. Damasio's (1996) Somatic Marker Hypothesis has been extremely influential in neuropsychological and psychophysiological studies of decision-making (Bechara & Damasio, 2005; Dunn et al., 2010; Suzuki et al., 2003; Verdejo-García & Bechara, 2009) and moral decision making (Koenigs et al., 2007; Moretto et al., 2010; Young et al., 2010). However, individual differences in interoceptive capacities have not previously been explored in the context of physiological arousal and moral decision making.

Again, moral judgments of harm were the focus of this study, due to the established emotional and physiological processes shown to influence harm-rejection judgments in sacrificial dilemmas (e.g. Carmona-Perera et al., 2013; Cushman et al., 2012; Greene et al., 2001; Moretto et al., 2010). Arousal is a critical aspect of emotional experience (Russell & Barrett, 1999), and people's sensitivity towards changes in arousal states in their bodies (e.g. Katkin, 1985) is likely to influence subjective experiences of emotion (e.g. Feldman Barrett et al., 2004; Wiens et al., 2000). Heartbeat detection ability in particular, has been shown to influence cardiovascular reactivity to aversive stimuli (Eichler & Katkin, 1994; Pollatos et

al., 2007b), facilitate adaptive regulation of emotion (Kever et al., 2015; Pinna & Edwards, 2020) and psychophysiological arousal (Füstös et al., 2013).

We measured interoceptive sensibility, accuracy, meta-cognitive awareness (Garfinkel & Critchley, 2013), and gastric interoception (Van Dyck et al., 2016) to understand whether these capacities were related to moral judgments in different ways. Physiological responses during a moral dilemma task were measured, to test whether interoceptive accuracy moderated the relationship between anticipatory physiological arousal prior to making moral judgments and the moral judgments themselves. Heartbeat detection ability was assessed using a heartbeat counting task (HBC; Schandry, 1981), as a relatively simple method of assessing people's awareness of discrete autonomic events. This task appeared to be more suitable to individual differences experiments (e.g. Dunn et al., 2010), than heartbeat discrimination tasks that require people to discern whether their heartbeats coincide with a tone (HBD; Whitehead & Drescher, 1980), largely due to the difficulty of the task (Kleckner et al., 2015). There is also a much greater task-load for participants, with evidence suggesting that a minimum of forty trials is required to achieve adequate reliability and statistical power (Kleckner et al., 2015). Furthermore, the HBD task is likely to capture people's ability to integrate internal and external sensory information whereas the HBC task requires attention to be completely focused on visceral sensations (Forkmann et al., 2016). This HBC method was considered more appropriate to study 2 as we were interested in how someone's habitual capacity to notice changes in physiological arousal may modulate the influence of arousal on decision-making.

Gastric interoception was measured using a two-step water load task (Van Dyck et al., 2016). Sensitivity to gastric processes and sensations of disgust have been shown to influence moral judgments (Schnall et al., 2008; Vicario et al., 2018), but this method has not previously been used in this context. Gastric sensitivity measures are also scarcely used in interoception research, and so this provided an opportunity to explore correlations previously found between cardiac and gastric forms of interoception (e.g. Herbert et al., 2012). The water-load test (WLT) was first used by Boeckxstaens et al (1990) as a non-invasive alternative to barostat measures of gastric sensitivity. The original test either asked participants to drink freely for a period of five minutes to maximum fullness or demanded water be ingested at a rate of 100ml per minute (De Schepper et al., 2004). Other water-load test methods have used a three-litre maximum water volume provided to participants across

two five-minute drinking intervals (Van Dyck et al, 2016). Earlier experiments have provided 3.5 litres of water during a ~30-minute water loading procedure (Schoen, 1957). To mitigate the small risk that participants suffer from hyponatremia (loss of sodium in the blood), older participants (60 +) were excluded (O'Neill, 1996), a small amount of electrolyte solution was added to the water, and the maximum amount of water provided to participants was only 1.3 litres for a period of five minutes as used in Zeng et al (2007). The two sensory thresholds we used were 1) noticing the first signs of stomach distention (Zeng et al, 2007) and 2) maximum fullness. Prior studies using the WLT in the context of eating-disorders have used 'satiety' as the first sensory threshold (Van Dyck et al., 2016). However, we were interested in people's sensitivity to the onset of visceral sensations of stomach filling up to maximum fullness, rather than experiences of satiety, to provide an indication of how early in the stomach filling process people notice signals associated with the stomach expanding.

As in Study 1, the moral dilemma stimuli allowed the calculation of harm-aversion and outcome-maximisation tendencies separately (Conway & Gawronski, 2013) which has not been possible for earlier psychophysiological studies using traditional moral dilemmas (e.g. Greene et al., 2001; Patil et al., 2014; Moretto et al., 2010). Furthermore, these stimuli provided a useful control condition for comparing the change of physiological arousal when contemplating moral 'dilemmas' that involved no conflict between harm-aversion and outcome-maximisation motivations i.e. in congruent conditions only harm-aversion drives the response, and dilemmas that did provide a conflict (incongruent dilemmas). Therefore, we could calculate the difference in physiological arousal associated specifically with this moral conflict element of the dilemmas, rather than for example, comparing the difference in physiological arousal in 'moral' dilemmas to 'non-moral' dilemmas (e.g. Greene et al., 2001). Furthermore, by using 'congruent' versus 'incongruent' dilemmas, we could specifically explore differences in physiological arousal associated with motivations of outcome-maximisation and harm-aversion. Congruent and incongruent dilemmas were also matched, which reduced the influence of confounding factors associated with the content of the text. Although, some of the dilemmas are more similarly matched in 'story' content than others.

As this study was largely exploratory, we captured a range of physiological parameters used in prior moral dilemma research, to test whether these distinct parameters showed relationships with moral judgments, and whether interoception interacted with these

parameters. We standardised the physiological data within-participants to account for any bias associated with using absolute values, which can vary considerably between people (Braithwaite & Watson, 2015). Electrodermal activity is a solely sympathetically mediated measure of physiological arousal that is characteristically absent during moral dilemma tasks involving interpersonal harm for people with damage to the ventromedial prefrontal cortex (Moretto et al., 2010). Heart rate is sympathetically and parasympathetically mediated (Glick & Braunwald, 1965) and has been associated with emotional experience (Lang et al., 1993) and the valence of emotional response towards moral dilemmas of interpersonal harm (Carmona-Perera et al., 2013). Finally, the threat-reactivity index used is a relative measure of systemic vascular resistance (implicated in action-based harm aversion: Cushman et al., 2012) and cardiac output, that is proposed to represent underlying motivational states of threat and challenge (Mendes et al., 2007). By gathering a more complete picture of changes to physiological arousal, we could be more precise in our conclusions about how distinct physiological responses may represent particular anticipatory emotional or motivational processes driving particular moral judgment response tendencies.

Finally, a study using the same moral judgment stimuli found that vagal-tone (an index of neurovisceral integration associated with superior emotional regulation) was associated with outcome-maximisation but not harm-aversion, and suggested that this is because harm-rejection judgments do not require neuro-visceral integration (Park et al., 2016). However, a recent study (Parton & McGinley, 2019) using a small set of moral dilemmas, did not replicate this finding, and suggested that vagal tone was not a reliable predictor of moral judgments. Like vagal tone, interoceptive accuracy has been associated with enhanced emotional regulation, particularly down-regulating negative emotional states (Füstös et al., 2013; Pinna & Edwards, 2020) which could influence the relationship between physiological arousal and outcome-maximisation and harm-aversion responses. Although anticipatory physiological arousal typically occurs for dilemmas involving direct interpersonal harm (Carmona-Perera et al., 2013; Cushman et al., 2012; Moretto et al., 2010), harm-aversion responses have been associated with both an aversion to harmful acts (action-aversion) and an aversion to the negative consequences of harm (outcome-aversion) (Reynolds & Conway, 2018). This suggests that harm-aversion tendencies can be influenced by negative emotional processes that are other-focused (outcome) and self-focused (action), which may reveal different associations between physiological processes and moral judgments. Conversely, recent research has challenged the premise of dual process models

(Greene et al., 2001) that ‘utilitarian’ judgments are less ‘intuitive’ or require more deliberation than ‘deontological’ judgments (Bago & De Neys, 2019). Therefore, we used the complete moral dilemma stimuli that includes both direct and indirect acts of harm (Conway & Gawronski, 2013) to understand whether an ability to accurately perceive cardiac sensations accounted for any differences in the link between physiological arousal and harm-aversion and outcome-maximisation response tendencies.

Statement of Authorship

This declaration concerns the article entitled:			
Do individual differences in interoception influence the relationship between physiological arousal and moral judgments of harm?			
Publication status (tick one) **Draft manuscript** [x] **Submitted** [ ] **In review** [ ] **Accepted** [ ] **Published** [ ]			
Publication details (reference)	Draft paper to be submitted to: *Cognitive, Affective and Behavioural Neuroscience* Authors: Helen Brown, Michael J. Proulx and Danaë Stanton Fraser		
Copyright status (tick the appropriate statement) I hold the copyright for this material [X] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed the formulation of ideas, design of methodology, experimental work and presentation of data in journal format. The candidate (H.B.) conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	This information has been redacted for privacy reasons.	**Date**	05/11/2020

Do individual differences in interoception influence the relationship between physiological arousal and moral judgments of harm?

Helen Brown [1,2], Michael J. Proulx [2] and Danaë Stanton Fraser [1]

[1] CREATE Lab, Department of Psychology, University of Bath, Bath, UK

[2] Crossmodal Cognition Lab, Department of Psychology, University of Bath, Bath, UK

Author note:

Research funded by University of Bath. This studentship runs alongside UK EPSRC grant Virtual Realities: Immersive Documentary Encounters (ref: EP/P025595/1).

Correspondence concerning this article should be addressed to Helen Brown, hrhb20@bath.ac.uk.

Abstract

The Somatic Marker Hypothesis (Damasio et al., 1990; Damasio, 1996) was among the first theoretical accounts to explain how the processing of somatic signals in the brain could bias decision-making concerning ethical violations. The role of emotional and physiological factors in judgments concerning the morality of harmful acts is now well-established (Cushman et al., 2012; Greene et al., 2001, 2004; Miller et al., 2014; Reynolds & Conway, 2018). Yet, individual differences in how people typically perceive or interpret physiological sensations in the body in the context of moral decision-making is little understood. Interoception broadly represents our sensory capacities to perceive visceral sensations in the body that are fundamentally connected with processes that manage homeostasis (Craig, 2015). Interoceptive processes are also closely tied to emotional experiences and motivation (Barrett & Simmons, 2015; Craig, 2015; Duschek et al., 2015). This within-participants experimental study (n=90) explored whether individual differences in interoception influenced the relationship between physiological arousal and moral judgments of harm. Interoceptive sensibility, interoceptive accuracy, interoceptive meta-cognitive awareness (Garfinkel et al., 2015) and gastric interoception (Van Dyck et al., 2016) were measured. Moral dilemma stimuli enabled the calculation of harm-aversion and outcome-maximisation tendencies independently (Conway & Gawronski, 2013). Cardiac impedance, heart rate and electrodermal activity were measured during a moral dilemma task. We found a conditional effect of interoceptive accuracy in the relationship between anticipatory sympathetic cardiac impedance, heart rate and harm-aversion tendencies, and between heart rate and outcome-maximisation tendencies. Interoceptive meta-cognitive awareness and gastric interoception also moderated the relationship between electrodermal autonomic arousal and judgment response time. We suggest interoceptive accuracy may support enhanced regulatory processes when individuals are faced with aversive moral dilemmas, that can attenuate or strengthen the relationship between indices of physiological arousal and moral judgments.

Keywords: interoception, moral dilemmas, moral judgment, outcome-maximisation, harm-aversion, harm, arousal, heartbeat detection, gastric interoception.

Introduction

When making moral decisions in situations where we feel conflicted or have limited information, we may default to what feels 'right' to us in the moment (Haidt, 2001; Tversky & Kahneman, 1974). Oftentimes, we cannot access the origin of these inclinations, but metaphors relating to 'gut feelings' or 'intuition' helpfully conceptualise impalpable visceral processes that guide us towards or away from certain moral choices. Damasio's Somatic Marker Hypothesis (SMH; Damasio et al., 1990; Damasio, 1996) was among the first theoretical accounts to explain how the processing of physiological experiences in the ventromedial prefrontal cortex (VMPfC) could influence judgments and behaviour concerning moral violations (Damasio et al., 1990; Koenigs et al., 2007; Moretto et al., 2010; Young et al., 2010). The SMH proposes that the processing of somatic signals generated within the body provide intuitive guidance during decision-making. Damage to the VMPfC has been linked to deficits in social emotions such as empathy and guilt (Anderson et al., 2013), anti-social behaviour (Damasio et al., 1990), and greater acceptance of personally harmful actions in moral dilemma paradigms (e.g. Moretto et al, 2010). In particular, individual differences in the aversive physiological response people have to the prospect of condoning personal harmful acts has shown to predict moral judgments harm (Carmona-Perera et al., 2013; Cushman et al., 2012; Greene et al., 2001, 2004; McDonald et al., 2017; Moretto et al., 2010). A physiological aversion to harmful acts is believed to arise due to learned associations between characteristically harmful acts and human suffering (the unconditioned aversive stimulus). Over time the harmful act itself, even if only imagined, is able to generate an aversive emotional and physiological response (Blair, 1995; Miller et al., 2014). However, people demonstrate distinct aversions to harmful actions and the negative outcomes of harm (Miller et al., 2014) which have both shown to be influential when forming moral judgments of harm (Reynolds & Conway, 2018).

Interoception

Our perceptual sensitivity towards visceral processes is also likely to be important in moral decision-making processes. For example, the sound of 'quickening' heartbeat feedback (Gu et al., 2013) has shown to bias ethical decision-making; demonstrating how the perception of physiological arousal alone is powerful enough to influence decision-making. Individual differences in sensitivity towards gastric sensations has been shown to predict the harshness of moral judgments (Schnall et al., 2008, 2015) and an awareness of cardiac

sensations has been associated with a propensity for antisocial behaviour (Nentjes et al., 2013). Interoception refers to our perceptual awareness of visceral processes concerned with maintaining the homeostatic state of the body, such as those happening within the cardiovascular or gastrointestinal systems (Craig, 2015). Interoceptive signals regarding the body's current state are periodically sent to the brain, which provides a motivational context (Damasio, 1996; Suzuki et al., 2003), colouring emotional experiences (Barrett et al., 2004) and facilitating the selection of adaptive behavioural responses (Bechara et al., 1997; Craig, 2015; Damasio, 1996). Even though emotional and physiological experiences appear to play a critical role in moral decision-making (Damasio et al., 1990; Greene et al., 2001, 2004; Moretto et al., 2010), interoceptive processes have been comparatively understudied in this field. The current study is the first to investigate the role of individual differences in interoception in the relationship between anticipatory physiological arousal and moral judgments of harm.

Interoception is a multidimensional construct, consisting of sub-components of interoceptive abilities that are not necessarily related (Forkmann et al., 2016; Garfinkel & Critchley, 2013). Most commonly studied is interoceptive accuracy (IAc), which typically measures individual differences in people's detection (Schandry, 1981) or differentiation (Whitehead et al., 1977) of visceral signals such as heartbeats. How much we might tend to focus on internal sensations such as heartbeats i.e. interoceptive sensibility (IS; Garfinkel & Critchley, 2013), is typically not correlated with interoceptive accuracy (Ainley & Tsakiris, 2013; Ferentzi et al., 2018b). A measure that combines conscious awareness of sensations with perceptual ability is interoceptive meta-cognitive awareness (Garfinkel et al., 2015). Confidence ratings of the accuracy of our visceral percepts (e.g. heartbeats), can provide an indication of an individual's awareness of the reliability of their interoceptive sensitivity. There is some evidence to suggest that cardiac and gastric interoceptive sensitivity may be related (Herbert et al., 2012; Whitehead & Drescher, 1980). Measures of gastric interoception are of particular relevance as physiological states, such as hunger have shown to influence moral judgments (Vicario et al., 2018), potentially due to hormonal changes associated with food consumption that increase feelings of nausea and potentially disgust, which may generalise to judgments of moral violations (Tracy et al., 2019). Consistent with this, individual differences in disgust sensitivity (Vicario et al., 2018) and how much people typically focus on gastric sensations (Schnall et al., 2008) can influence moral judgments, further highlighting the embodied nature of moral decision-making.

Individual differences in interoception may also moderate the relationship between physiological arousal and moral decision-making due to how people habitually respond to surprising interoceptive sensations. Predictive-coding models of interoception propose that prediction error in the brain represents a deviation from an expected interoceptive state, which indicates how 'surprising' interoceptive sensations are, within a given context (Seth, 2013; Seth & Friston, 2016). More surprising sensations associated with greater prediction errors, are more likely to reach conscious awareness. Prediction error can lead to active and perceptual inference (Apps & Tsakiris, 2014; Farb et al., 2015; Seth & Friston, 2016). With active inference, the brain aims to match the interoceptive state with the expected state, facilitating behavioural responses to achieve sensory regulation. With perceptual inference, the expected state is altered to be in sync with the interoceptive state (Seth & Friston, 2016), meaning we are less fixed to prior interoceptive expectations and optimising the precision of the current sensation is prioritised (Farb et al., 2015). Farb and colleagues (2015) suggest more accurate interoceptive representations would be associated with perceptual inference, because with active inference 'error' states are more likely to be maintained. Farb et al (2015) also propose that perceptual inference may reduce overt regulatory behaviours directing at resolving prediction error, if a closer examination of the origin and nature of the interoceptive signal is allowed.

An enhanced ability to direct attention towards heartbeat sensations is likely to be associated with enhanced precision of heartbeat signals (Ainley et al., 2016) which may support superior regulation of psychophysiological arousal states (e.g. Füstös et al., 2013). Consistent with this notion, interoceptive accuracy (IAc) has been associated with enhanced self-regulatory abilities during exercise (Pollatos et al., 2007a), pain tolerance (Weiss et al., 2014) and antecedent and response-focused emotional regulation strategies (Füstös et al., 2013; Kever et al., 2015). Heartbeat detection ability has been associated with more intense emotional experiences (Feldman Barrett et al., 2004; Wiens et al., 2000) and enhanced physiological reactivity to emotional stimuli and physical stressors (Eichler & Katkin, 1994; Pollatos et al., 2007b) suggesting a strong association between physiological arousal and subjective emotion in people who are more sensitive to visceral sensations. Importantly, trait anxiety has also been associated with IAc (Critchley et al., 2004; Pollatos et al., 2007a) and greater reactivity to emotional stimuli (Takahashi et al., 2005). Habitually higher levels of autonomic arousal may therefore support heartbeat detection ability (Eichler et al., 1987; Eichler & Katkin, Edward, 1994).

However, anxiety has also been linked to a hyper-awareness towards visceral sensations (Anderson & Hope, 2009; De Berardis et al., 2007) which may reflect a maladaptive self-reported tendency to focus on internal sensations along the dimension of interoceptive sensibility (IS) (Mehling et al., 2018). People who pay attention to bodily sensations more often are more likely to misinterpret changes in physiological sensations as meaningful (Clark et al., 1997; Domschke et al., 2010), which can intensify these experiences and increase expectations of negative outcomes and anxiety (Paulus & Stein, 2010). When bodily feedback is available, IS has shown to predict risk-averse behaviour (Salvato et al., 2019). IS has also been indirectly associated with a reduced aversion to harmful acts, and more intuitive and incorrect responses on the Cognitive Reflection Task (Brown et al., 2020) which is proposed to provide a measure of people's capacity to 'override' an intuitive response to a series of counterintuitive puzzles (Frederick, 2005). High IS may predict a reliance on intuitive heuristics when faced with cognitively demanding tasks like moral dilemmas (Brown et al., 2020).

Moral dilemmas

Historically, experimental moral dilemma paradigms exploring moral judgments of harm used conflict-dilemmas, such as the Trolley (Thomson, 1985) and Footbridge (Foot, 2003) problems. These dilemmas ask people whether it is appropriate to sacrifice one person to save several more people on a train track in the path of a runaway train. In the Trolley dilemma, this choice is made by pulling a hypothetical lever to divert the path of the oncoming train to another track that has one person on it. In the Footbridge dilemma, the choice is made by pushing a large man of a bridge into the path of the runaway train, thus stopping the train. Although the outcome of choosing to harm the one person is the same in both dilemmas, people are more likely to accept harm in the Trolley dilemma and reject harm in the Footbridge dilemma (Greene et al., 2001). People's inclinations to reject or accept harm, have typically been conceptualised as 'utilitarian' or 'deontological', although recent work has shown these judgments do not necessarily represent the proposed underlying moral philosophies (Kahane et al., 2018). Utilitarianism refers to the moral framework that good actions are those that maximise the wellbeing for the greatest number of people (Mill, 1998). Whereas, deontological ethics evaluates the goodness of an action on how the act itself upholds the rights and duties of individuals within a situation (Kant, 2018). Causing harm in both the Trolley and Footbridge dilemma is seen as morally appropriate in a utilitarian sense,

if it maximises outcomes, whereas a deontological stance would see harm as morally wrong regardless of the outcomes, as the act itself disregards the rights of the person being harmed.

Moral judgments can be influenced by an aversion to the harmful action itself (action-aversion), and an aversion to the negative outcomes of the harmful action (outcome-aversion) (Cushman, 2013; Miller et al., 2014). Furthermore, a motivation to reject harm and a motivation to maximise outcomes do not represent two ends of a bipolar scale (Conway & Gawronski, 2013). Traditional conflict dilemmas are unable to differentiate people's motivation to reject harm ('deontological') or maximise outcomes ('utilitarian'), as they only measure the strength of these competing motivations to inform the final moral judgment (Conway & Gawronski, 2013). A more recent method developed by Conway and Gawronski (2013) using process-dissociation (Jacoby, 1991), makes it possible to independently calculate people's harm-aversion, and outcome-maximisation response tendencies. Participants make judgments in paired 'incongruent' and 'congruent' dilemmas that are matched in story structure but vary in whether harmful action maximises outcomes. Incongruent dilemmas are the same as traditional moral dilemmas (i.e. causing harm results in a greater good), whereby utilitarianism and deontological ethics disagree on the acceptability of harm. Here, individual differences in motivations to avoid harm and maximise outcomes would lead to divergent responses. Whereas in the congruent dilemmas, no greater good is achieved by causing harm, therefore utilitarianism and deontological ethics are in agreement that harm is wrong. Using a probability equation (see Conway & Gawronski, 2013), it is then possible to calculate someone's tendency to maximise outcomes accounting for their tendency to reject harm.

Physiological responses to harm-based moral dilemmas

Harm rejection judgments have been associated with automatic, emotional and physiological processes in traditional moral dilemmas (Cushman et al., 2012; Greene et al., 2001; McDonald et al., 2017; Moretto et al., 2010). Greene and colleagues influential dual process model (Greene et al., 2001, 2004) proposes that powerful emotional responses to the prospect of harming the one person, or insufficient resources to engage in cognitive deliberation to override this response, would predict strong harm-rejection judgments. Some recent work has challenged the notion that utilitarian decisions are less intuitive than deontological decisions (Bago & De Neys, 2019; Kahane, 2014). However, there is considerable empirical evidence favouring the dual process account. Utilitarian or outcome-

based choices are associated with more deliberative cognitive processes (Baron et al., 2015; Greene et al., 2004; Patil et al., 2020), stress-inductions reduce utilitarian judgments and increase response time (Starcke et al., 2012), and an inability to identify emotional states has been associated with stronger utilitarian judgments (Patil & Silani, 2014). Physiological indices of sympathetic autonomic arousal believed to underlie motivational threat states (Blascovich & Mendes, 2000), such as increases in systemic vascular resistance (SVR; tightening of blood vessels following the release of epinephrine) (Cushman et al., 2012), and shortening of pre-ejection period (PEP; stronger myocardial contractility) (Parton & McGinley, 2019) have been associated with action-based harm aversion. Individual differences in cardiovascular threat reactivity when carrying out simulated harmful acts, have been shown to predict harm-rejection responses to sacrificial moral dilemmas when using a measure of SVR (Cushman et al., 2012), but not when using a measure of PEP (Parton & McGinley, 2019). Increases in skin conductance (McDonald et al., 2017) and anticipatory heart rate acceleration (Francis et al., 2016; Starcke et al., 2012) have been associated with an aversion to direct interpersonal harm. However, heart rate deceleration is typically associated with negative emotional experiences to aversive stimuli (Lang et al., 1993; Palomba et al., 2000) including moral dilemmas involving direct interpersonal harm (Carmona-Perera et al., 2013). Finally, lower vagal tone (resting heart rate variability) has shown to predict strong outcome-based moral preferences (Park et al., 2016), which the authors suggested is due to a reduced ability to integrate physiological signals into judgment formation; although this finding was not replicated in a later study (Parton & McGinley, 2019).

Interoception and moral judgment

It is possible that individuals higher in interoceptive accuracy (IAc) may habitually use forms of perceptual inference when faced with interoceptive prediction error (Ainley et al., 2016; Farb et al., 2015). Better heartbeat detection ability has shown to facilitate the modulation of affect-related arousal (Füstös et al., 2013) and enhanced emotional regulation processes (Kever et al., 2015), which, in the context of moral decision-making, could attenuate the relationship between anticipatory physiological arousal and harm-rejection judgments. If changes in physiological arousal lead to active inference (purportedly associated with less accurate interoceptive representations; Farb et al., 2015), this could encourage ‘knee-jerk’ responses to reject harmful actions, as a regulatory process aimed at alleviating an aversive arousal state.

An alternative hypothesis is that interoceptive accuracy strengthens the relationship between physiological arousal and moral judgments as better heartbeat detectors have demonstrated enhanced psychophysiological reactivity towards emotional stimuli (Eichler & Katkin, 1994; Pollatos et al., 2007b). Furthermore, using an intuitive reasoning task (IRT), Dunn et al (2010) found when anticipatory somatic signals supported advantageous intuitive decisions, enhanced IAc was helpful, whereas when somatic signals favoured disadvantageous decisions, IAc was unhelpful (see also Werner et al., 2009). Greater accessibility of interoceptive sensations could make higher IAc people more susceptible to concurrent physiological states when making moral judgments. However, although people may be conflicted about moral judgments, the nature of moral dilemmas means they do not (typically) have a 'right' answer, and the utility of somatic signals for predicting negative outcomes of moral decisions will not be positively or negatively reinforced during a moral dilemma task - unlike the IRT task (Dunn et al., 2010). Therefore, this moderation effect of IAc may manifest differently in the relationship between anticipatory arousal and moral judgment.

Finally, previous research (Brown et al., 2020) suggests that interoception may influence egocentric harm judgments e.g. '*Would you harm…*?', in a different way to allocentric harm judgments e.g. '*How acceptable do you find…harmful action*?'. Egocentric moral judgments require a degree of self-referential processing as we place ourselves in the shoes of the harmful actor to imagine self-relevant consequences (Sood & Forehand, 2005); which appear to engage emotional processes that allocentric judgments do not (Berthoz et al., 2006; Tassy et al., 2013). Therefore, as interoception and egocentric judgments both entail a level of self-referential processing, interoception may be more influential for these types of judgments.

Present study

This study investigated the role of interoception in moral judgments of harm with a series of pre-registered (Brown et al., 2019; https://osf.io/m6f5j) and exploratory research questions. Physiological measures included electrodermal responses and cardiac impedance to capture anticipatory changes in sympathetic arousal previously associated with the rejection of personally harmful acts (e.g. Cushman et al., 2012; Moretto et al., 2010). Heart rate was also measured as a correlate of negative affective experiences (Lang et al., 1993), which has been associated with moral dilemmas involving direct interpersonal harm

(Carmona-Perera et al., 2013). We explored whether interoception of the cardiac and gastrointestinal system predicted moral appropriateness (egocentric) and moral acceptability (allocentric) judgments of harm, and whether interoceptive accuracy moderated the relationship between physiological arousal and moral judgments. As these specific physiological indices have not previously been measured using this moral dilemma paradigm, we also explored whether physiological arousal predicted outcome-maximisation or harm-aversion tendencies. An exploratory research question was whether interoception moderated the relationship between physiological arousal and judgment response time.

Pre-registered research questions

R0. Does BMI, age, sex, time estimation accuracy or anxiety correlate with interoceptive accuracy or gastric sensitivity?

R1. Is interoceptive accuracy associated with gastric interoception?

R2. Does physiological arousal predict moral judgments of harm?

R3. Does interoception predict moral judgments of harm?

R4. Does interoceptive accuracy moderate the relationship between physiological arousal and moral judgments? *Note*: only interoceptive accuracy was included in R4, which is a refinement of the pre-registered research question that broadly stated 'interoception'.

Exploratory research question

ER1. Does physiological arousal predict response-time differences between congruent and incongruent dilemmas, and is this moderated by interoception?

Method

Design

A within-participants experimental design was used. Moral appropriateness and moral acceptability judgments, and physiological responses were the primary dependent variables. Interoception and physiological responses were the primary independent variables. The

potential influence of age, sex, time-estimation accuracy, BMI and anxiety on interoception measures was investigated.

Participants

Following approval from the University of Bath ethics committee, participants were recruited from the University of Bath Student Participant Pool, the Psychology Department community panel database and via departmental email lists. Participants received course-credit or £15 in exchange for their time. The study was advertised on the University noticeboard in the psychology department and various online recruitment platforms. Minimum sample size (n=77) was based on a priori power calculations using G*Power (α = 0.05, β = 0.8, f^2=.015, df= 3, 71). A total sample of 90 participants took part in the study, comparable to similar interoception studies (e.g. Van Dyck et al, 2016; Dunn et al., 2010). Healthy right-handed participants aged 18-60 were recruited. Participants were not eligible to take part if they: were pregnant; engaged in intense physical activity regularly; had a heart/gastrointestinal condition or received surgery to those areas; had a history of mental illness; history of renal problems; had current physical conditions or medication that affect diet or weight or ability to safely take part in the study. The sample was 77.8% female and there was an age range of 18-50 years old (Median=21, SD=6.35). Two participants were left-handed but completed the computer task with their right hand.

Questionnaires

Anxiety

The State and Trait Anxiety Scale (Spielberger & Gorsuch, 1983) was used to measure anxiety. The scale included two identical 20-item scales with positive and negatively coded items. Participants reported their agreement (Not at all/Somewhat/Moderately so/Very much so), with twenty different statements e.g. *'I feel calm', 'I feel tense'*. The State and Trait measures ask people to rate how they feel right now and how they feel in general, respectively. Cumulative scores representing State and Trait anxiety were calculated.

Interoceptive sensibility

Interoceptive sensibility was measured using the 'Private Body consciousness' subscale of The Body Consciousness Questionnaire (BCQ: Miller et al., 1981) which captures people's tendency to notice bodily sensations, and has been used in prior interoception research (e.g. Ainley & Tsakiris, 2013; Sze et al., 2010; Werner, Duschek, et al., 2009). The subscale includes 5 statements including: *'I'm very aware of changes in my body temperature', 'I am quick to sense the hunger contractions in my stomach', 'I know immediately when my mouth or throat gets dry', 'I am sensitive to internal bodily tensions',* and '*I can often feel my heart beating'*. Participants reported how characteristic they believed each statement was (extremely uncharacteristic/uncharacteristic/ neutral/characteristic/extremely characteristic). The complete BCQ was used to maintain scale-validity. Scores for the Private Body Consciousness subscale were numerically coded 1-5 (1=extremely uncharacteristic) and a mean score calculated.

Moral dilemmas

Moral dilemma stimuli developed by Conway and Gawronski (2013) were implemented to calculate harm-aversion and outcome-maximisation tendencies when faced with moral dilemmas of harm. Outcome-maximisation refers to the tendency to condone harmful action that results in a greater benefit or good overall. Harm-aversion represents the tendency to reject harmful actions, irrespective of the consequences of harm. The original stimuli include 20 moral dilemmas comprising 10 story-pairs. Story-pairs are matched in their content, but whereas one dilemma asks people whether it is morally appropriate to carry out harmful action that maximises outcomes (incongruent), the other proposes a harmful action that does not maximise outcomes (congruent); see Figure 1 for examples. People provide a judgment about whether they condone the harmful action or not: 'Yes, this is appropriate' or 'No, this is not appropriate'. In the incongruent trials, people's motivations to avoid harm or maximise outcomes could both be driving their moral judgment. Whereas, in the congruent trials, only an aversion harm will drive decision-making as there is no conflicting incentive to maximise outcomes (see Conway & Gawronski, 2013, for full probability equations). Harm-aversion and outcome-maximisation response tendencies were calculated based on the frequency that harm was accepted and rejected in congruent and

incongruent trials. These raw scores were subsequently transformed into z-scores. Dilemmas were presented in the order suggested by Conway and Gawronski (2013).

Figure 1.

Example of pair of moral dilemma stimuli (Conway & Gawronski, 2013)

Incongruent: *Causing harm maximises outcomes*	**Congruent:** *Causing harm does NOT maximise outcomes*
You find a time machine and travel back to the year 1920. While checking into a hotel, you meet a young Austrian artist and veteran of the First World War. You realize this is Adolf Hitler before his rise to power in Nazi Germany. He is staying in the hotel room next to yours and the doors are not locked. It would be easy to simply smother him with a pillow in his sleep and disappear, stopping the Second World War and the Nazi party before they even start. However, he has not committed any crimes yet and it seems wrong to hurt an innocent person. **Question:** Is it appropriate for you to kill an innocent young Hitler in order to prevent the Second World War?	You find a time machine and travel back to the year 1920. While checking into a hotel, you meet a young petty criminal. You realize this is George Brackman, a man who later on abducted a child and held her for a week until her family paid him some ransom money. He is staying in the hotel room next to yours and the doors are not locked. It would be easy to simply smother him with a pillow in his sleep and disappear, stopping the abduction and ransom demands before they even start. However, he has not committed any crimes yet and it seems wrong to hurt an innocent person. **Question:** Is it appropriate for you to kill George Brackman in order to prevent him from taking a child hostage?

Moral acceptability

In addition to the harm aversion and outcome-maximisation measures, we implemented a moral acceptability scale. This allowed us to assess the strength of people's moral judgments of utilitarian and non-utilitarian harmful acts proposed in the moral dilemmas from an allocentric perspective. The item read: *'How morally acceptable or morally unacceptable do you find the proposed action to be?*', with response options: 1 = Completely unacceptable, 2 = Moderately unacceptable, 3 = Slightly unacceptable, 4 = neither acceptable nor unacceptable, 5= Slightly acceptable, 6= Moderately acceptable, 7= Completely acceptable (adapted from Schnall et al., 2008).

Interoception tasks

Interoceptive accuracy and meta-cognitive awareness

Cardiac interoceptive accuracy was measured using procedures from Schandry's (1981) heartbeat detection task (HBD). Participants were seated, asked to remove any watches, and a pulse oximeter (BLYL CMS 50D+) was attached to the index finger on the right hand. They were instructed to silently count their heartbeats for four trial periods; 30, 50, 40, and 20 seconds (in the same order). The experimenter signalled the beginning of each trial and the end of each trial was signalled by an alarm. For each trial participants reported how many heartbeats they perceived and verbally reported how sure they were that their answer reflected the actual amount of heartbeats; *How sure are you that the amount of heartbeats you counted for that interval were correct?* (0=not sure at all, 8= absolutely sure). Participants underwent a time estimation task to test whether time accuracy ability correlated with performance on the HBD task. Equivalent to the HBD task, participants were asked to estimate the length in seconds of four trial periods of 45, 35, 55, and 25 seconds. Within-subjects' correlations (Pearson's *R*) of tracking accuracy and confidence ratings provided the interoceptive meta-cognitive awareness measure, representing participants' awareness of their own tracking ability, therefore, scores could range between 1 (perfect positive correlation) and – 1 (perfect negative correlation) and zero indicating no correlation.

The formula used to calculate interoceptive accuracy (and time accuracy) was adopted from Garfinkel et al (2015) and Hart (2013):

$$1- \frac{| nbeats_{real} - nbeats_{reported} |}{(nbeats_{real} + nbeats_{reported})/2} ;$$

Using this formula, scores can range from -1 to +1 (see Forkmann et al, 2016 for a discussion), where 1 is perfect accuracy. Moderate deviations between real heartbeats and reported heartbeats range between 0 and 1, whereas for larger differences between real and reported heartbeats "e.g., perceived <33% or >300% of recorded heartbeats" (pg 73, Forkmann et al, 2016), it is possible to obtain a negative value up to -1 (Forkmann et al, 2016; Hart et al, 2013). By having the reported heartbeats included in the denominator (whereas in Shandry's (1981) formula only real heartbeats are the denominator), this accounts for overestimations and underestimations in heartbeats. In a comparison of Shandry's (1981) and Hart's (2013) formulae for calculating heartbeat detection accuracy, Forkmann et al (2016) did not find evidence for a large influence of formula on interoceptive accuracy scores.

Gastric interoception

Gastric interoception was measured using a two-step Water-Load Test (Van Dyck et al., 2016) that was designed to measure interoceptive sensitivity to gastric processes. Participants sat in a normal upright position and drank non-carbonated tap water at room temperature. Participants were asked to drink water freely and as continuously as possible until they noticed 1) the first sensations of stomach bloating or distention (Zeng et al., 2007), and 2) when their stomachs were completely full of water. Participants notified the experimenter at each time point and the water consumed was measured. Participants were reminded they could stop the experiment at any time and to notify the experimenter should they feel unwell. Total drinking time was around 5 minutes. Participants drank from a 400ml opaque flask with a straw and provided with more water as needed in 200ml amounts to blind them to the amount they were drinking. A safe maximum amount of water available to participants was 1300ml, a maximum that they were not aware off. We added 2.5ml of electrolyte solution (Elete electrolyte) to the water to mitigate against a loss of sodium in the blood. Gastric interoception was calculated as the percentage of water drunk from the first

signs of stomach distension at step 1, to achieve maximum fullness at step 2, providing an index of gastric sensitivity not confounded by stomach capacity (Van Dyck et al., 2016). Gastric interoception thus represents the onset of somatic awareness from the first signs of stomach distension to complete fullness. We conceptualised lower scores as indicating an increased sensitivity to gastric processes, as fullness is reached after a longer period of drinking from first noticing stomach filling. As a manipulation check, a short hunger and thirst questionnaire was completed before and after the task: *How hungry do you feel at this moment?* (1=not at all, 9=extremely hungry), *How thirsty do you feel at this moment?* (1=not at all, 9=extremely hungry).

Physiological data collection

Impedance cardiographic (ICG), electrocardiographic (ECG) and galvanic skin-response signals (AUC) were recorded using a wireless BIOPAC MP160 (BIOPAC Systems, 2020) at a rate of 2000 samples per channel of data. Cardiac output (CO; volume of blood ejected from the heart at systole), systemic vascular resistance (SVR; amount of resistance in circulatory system to achieve blood flow around the body), heart rate (HR; beats-per-minute), and skin conductance activity (SC) were measured. Physiological waveforms were digitised, stored and analysed using *Acqknowledge* software (BIOPAC Systems, 2015). Electrodes were placed either side of participants' neck, and upper torso to capture ICG data. An Einthoven lead II configuration was used to collect ECG data. Electrodes on participants index and second finger captured electrodermal data. The wireless ICG, ECG and SC amplifiers were attached to participants with a chest-strap and wrist-strap. An estimation of SVR, CO and HR were calculated using Acknowledge software. A clinically validated blood pressure monitor (OMRON) was used to measure mean arterial pressure (MAP) (Cushman et al., 2012) which is used to estimate SVR. Blood pressure was measured twice during the experiment, specifically during presentation of 1) several congruent trials in a row and 2) several incongruent trials in a row. MAP was calculated using the formula:

$$MAP = (systolic\ pressure + 2(diastolic\ pressure))/3$$

Procedure

Participants were asked to not consume any food or drink for at least 2 hours or consume any alcohol or any (unprescribed) drugs or engage in intense physical exercise 24

hours prior to the study. After providing informed consent, participants completed all questionnaire measures, excluding the hunger/thirst scales. Participants provided some basic demographic details and completed a short health form to ensure they were feeling well and had completed the pre-study protocol. Their height and weight measurements were taken before completing the heartbeat detection and time-estimation tasks. Physiological data recording equipment was setup 5-10 minutes before recording. To aid waveform-calibration and allow physiological responses to stabilise at rest, data acquisition began 2-3 minutes before the moral dilemma task while participants were seated at the computer. E-Prime stimulus presentation software was used for the moral dilemma task, and instructions provided at the beginning. Participants were asked to keep as still as possible, uncross their legs and place their right arm resting near the keyboard. Blood pressure was taken before the experiment to habituate participants to the device and collect a baseline comparison. Their left arm remained raised and resting on a table in front of them (palm-up) throughout the moral dilemma task. Blood-pressure readings were taken at a further two time-intervals listed above. Presentation of moral dilemma stories was standardised depending on length of the text (*Range*= 21-35 sec; *mean* congruent = 28.89sec, *mean* incongruent = 29sec). Several participants reported missing key information during the allocated reading time, and the experimenter verbally repeated critical details missed in the text. Following each dilemma, participants were asked whether they condoned carrying out a harmful act associated with the dilemma and how morally acceptable they judged the *proposed* harmful action to be. They responded with a key press using their right hand. Response time was not capped to avoid time-pressured responses. A blank screen was presented for 3-seconds between the second moral judgment response and the presentation of the next moral dilemma text. After a short rest and removal of recording equipment, participants completed the water-load task and hunger/thirst scales. The experiment took approximately 1.5 hours to complete. Participants were debriefed and thanked for their time.

Physiological data processing

A 12.5msec delay was added to digital event marker channels before data processing to account for signal latency from the wireless amplifiers. Waveforms were checked and cleaned for motion-artefacts and noise. As ICG data is prone to motion-artefacts, a relaxed SFLC motion-artefact filter was applied to the majority of ICG waveforms. ICG waveforms for two participants required a more aggressive motion-artefact filter. Two SVR and CO

scores were calculated for each participant by inputting the associated MAP value for the congruent and incongruent trials into the ICG analysis in *Acqknowledge* (BIOPAC Systems, 2015). Average MAP was 82.04 ± 7.33 for congruent trials and 83.61 ± 7.89 for incongruent trials. Digital event markers symbolising presentation of the moral dilemma stories, questions and response-clicks were recorded on the data file to allow precise measurement-windows for analysis. Programming scripts were used to automate the calculation of mean CO, SVR, HR and SC data for the congruent and incongruent trials separately for the three measurement windows: 1) story contemplation phase; 2) between presentation of first moral judgment question and response; 3) between presentation of the second moral judgment question and response. A 1-second offset was applied to the beginning and end of SC measurement windows to account for delay in SC responses.

Cardiac reactivity

Due to the large inter-individual variation in physiological responses, all raw physiological data was transformed into relative values (*z*-scores) within each subjects' dataset to improve reliability and validity of individual differences analyses (Boucsein et al., 2012; Braithwaite et al., 2013; Braithwaite & Watson, 2015). Means and standard deviations used to derive *z*-scores were calculated for measurement windows 1-3. The CO and SVR scores were transformed into a combined threat-challenge reactivity score (see Scheepers et al., 2012; Turner et al., 2013). According to the biopsychosocial model (Blascovich & Mendes, 2010; Mendes et al., 2007), cardiovascular states are the product of an evaluation of whether the demands of the situation exceed our personal resources. When resources meet or surpass demands, a challenge motivational state occurs resulting in an efficient cardiovascular response associated with an increase in CO, and low SVR. A threat motivational state arises when the demands of the situation outweigh our resources, resulting in a reduction in CO and increase in SVR. To calculate the threat-challenge reactivity index, raw SVR scores were subtracted from CO scores. Higher scores reflect a higher challenge state, and lower scores indicate a higher threat state. Threat-challenge reactivity (TCR) and heart rate (HR) values for all trials within each timeframe (1-3 separately) were transformed into *z*-scores, and then *t*-scores to provide a relative measure of HR and TCR across all congruent and incongruent trials. Average *t*-scores were then calculated for each timeframe within congruent and incongruent trials.

Skin conductance

Skin conductance (SC) data was converted into Area Under the Curve (AUC; Naqvi & Bechara, 2006). A moving-average smoothing filter was first applied to the SC waveform (2000 samples). The waveform was then resampled at 125Hz, and a difference transformation with an interval of 6 samples was applied. Area under the curve is calculated as the area bounded by the curve and the chord connecting the juncture of the curve with the endpoints of each measurement window (Naqvi & Bechara, 2006). Area was divided by the time in seconds for that measurement interval, to calculate a rate of micro-siemens per second (AUC). This method provides an indication of both the amplitude and temporal features of electrodermal activity and does not require subjective interpretation about SC responses (Naqvi & Bechara, 2006). Within each measurement window (1) story contemplation, 2) pre-moral judgment 1 and 3) pre-moral judgment 2), a '*log (AUC + 1)*' transformation was applied to raw AUC scores to normalise the data while allowing for zero values (Braithwaite et al., 2013; Braithwaite & Watson, 2015). The logged AUC scores were then transformed into *z*-scores, and finally *t*-scores to remove negative values (see Braithwaite & Watson, 2015). Means and standard deviations used for these calculations were based on all trials (congruent and incongruent) within each measurement window. Average *t*-scores for measurement windows 1-3 (including those with zero values) were calculated for the incongruent and congruent trials separately, providing a time-corrected measure of AUC for each measurement window within congruent and incongruent trials.

Results

Statistical analysis was carried out using SPSS v26. Boxplots of the dependent variables were inspected for outliers. Data points with standardised residuals ± 3 standard deviations were investigated with boxplots, with any exclusions noted below. Pearson's bivariate correlations were carried out to assess relationships between independent and dependent variables. Data are mean ± standard deviation (SD) unless otherwise stated.

Moral judgments

Responses for one of the moral dilemma questions were not collected for all participants, due to a technical issue. To ensure consistent physiological and moral judgment data across the entire sample we excluded this moral dilemma pair and corresponding

physiological data from analysis, resulting in a total of 18 complete matched moral dilemmas: 9 congruent items and 9 incongruent items. Conway and Gawronski's (2013) formulae to calculate harm-aversion and outcome-maximisation parameters were adjusted accordingly. Eight participants were excluded from all moral acceptability judgment analyses due to a technical problem resulting in missing data on two trials. Boxplots revealed two outliers for the acceptability_congruent dependent variable who were also excluded from the moral acceptability analyses. A log10 transformation was carried out on the acceptability_congruent scores (2.873 ± .734) to normalise a strong positive skew. A Shapiro-Wilks test (SWT) showed the logged acceptability_congruent scores and acceptability_congruent (3.815 ± .812) scores had normal distribution (p>.05). The z-transformed outcome-maximisation and harm-aversion parameters were not normally distributed (SWT, p<.05) but inspection of Q-Q plots indicated fairly normal distribution. Harm-aversion and outcome-maximisation parameters showed non-significant negative correlation (r=-.143, p=.179), providing support for the distinction between these two motivations. Potential confounding variables of age, sex and anxiety did not correlate with outcome-maximisation or harm-aversion tendencies (p>.05), although men were more likely to judge harmful action as more morally acceptable on incongruent trials (r= -.236, p=.035).

Moral judgments across incongruent and congruent trials

Paired-sampled t-tests (bootstrappedx1000) were performed to assess whether participants were more likely to accept harmful action in the congruent dilemmas versus the incongruent dilemmas by comparing the percentage of accepting harm (i.e. 'Yes, this is morally appropriate') in congruent versus incongruent dilemmas. We found people condoned harmful action significantly more often for dilemmas where harm maximised outcomes, than when harm did not maximise outcomes (.309 ± .164), t (89) =17.87, p<.001, 95% CI [.274 to .342]. A second paired-samples t-test assessed whether people were more likely to accept or reject harm when harm maximised outcomes- this is the traditional 'utilitarian' versus 'deontological' analysis. Although people were marginally more likely to accept than reject harm in the interests of the 'greater good' (.067 ± .338), this was just below significance, t *(89)* =1.869, p=.051, 95% CI [-.002 to .128]. Finally, we tested whether participants rated harmful actions in the incongruent trials (where harm maximised outcomes) as more morally acceptable than in the congruent trials. Indeed, people judged harmful acts as more morally

acceptable on trials where harm maximised outcomes (3.374 ±.730), and this was significant, *t (79)* = 41.309, *p*=.001, 95% CI [3.219 to 3.538].

R0. Association between interoception and potential confounding variables

We explored bivariate correlations of all the interoception variables and potential confounding factors, including age, sex, BMI, anxiety and time-estimation accuracy (Table 1). Four participants were excluded from all interoception analyses: one participant disclosed they had guessed on the HBD task, and three did not follow the pre-study requirements regarding the consumption of food/drink. For the water-load task 8.1% of participants drank the maximum (1300ml) water available (mean=707.62, SD=305.14). We treated interoceptive accuracy as a continuous variable because the lower interoceptive accuracy group (using median-split procedure) showed a strong negative skew. None of the interoception variables showed correlation with each other, providing justification for conducting separate statistical analyses for each dimension. Age was significantly negatively correlated with interoceptive accuracy, suggesting younger people were more accurate in tracking heartbeats. Surprisingly, trait anxiety was negatively associated with IAc, suggesting higher levels of anxiety in general hindered heartbeat counting ability. Descriptive statistics of the interoception variables can be found in Table 2.

Table 1.

Bivariate correlations between Interoceptive Accuracy, Interoceptive Sensibility, Interoceptive meta-cognitive awareness, Gastric interoception and potential confounding variables

Variables	1	2	3	4	5	6	7	8	9	10
1. Age	1	-.136	.316**	-.090	-.016	-.010	-0.042	-.242*	.034	.191
2. Sex		1	-.091	.163	.256*	.168	-.027	-.183	-.090	.173
3. BMI			1	.144	.142	.006	.085	-.139	.002	.181
4. State Anxiety				1	.537**	.037	-.226*	-.120	-.152	.173
5. Trait Anxiety					1	.086	-.225	-.280**	-.188	.129
6. IS						1	.059	.028	-.049	-.048
7. Time Acc							1	.100	.002	-.241*
8. IAc								1	.015	-.176
9. IAc-meta									1	-.092
10. Int-Gastric										1

Note. 2. *p<.05, **p<.01. Body Mass Index= BMI, Interoceptive sensibility= IS, Time estimation accuracy= Time Acc, Interoceptive Accuracy = IAc, Interoceptive meta-cognitive awareness = IAc-meta, Gastric interoception = Int-Gastric.

Table 2.

Descriptive statistics for interoception variables

	Range	Minimum	Maximum	Mean	Std. Deviation
Gastric interoception	96.15	3.85	100	47.25	19.42
Interoceptive sensibility (IS)	3.60	.40	4.00	2.54	.77
Interoceptive accuracy (IAc)	1.98	-1.00	.98	.35	.52
IAc-meta cognitive awareness	1.97	-1.00	.97	-.07	.62

R1. Association between interoceptive accuracy and gastric interoception

A bootstrapped (x1000) linear regression model tested whether IAc was associated with gastric interoception. Interoceptive accuracy did not predict gastric interoception, R^2=.040, F (1, 84) =3.538, p=.063. IAc also did not predict the total amount of water drunk, R^2=.006, F (1, 84) =.526, p=.470, the amount of water drunk to achieve perceptual awareness of gastric distension, R^2=.003, F (1, 84) =.256, p=.615, or maximum fullness, R^2=.022, F (1, 84) =1.932, p=.168, on the task.

R2. Physiological arousal and moral judgments of harm

R2.1. Arousal differences between congruent and incongruent trials

Using paired-samples t-tests (bootstrappedx1000), we investigated whether physiological arousal was significantly different when participants contemplated dilemmas where harm maximised outcomes (incongruent) compared to dilemmas where harm did not maximise outcomes (congruent) (Table 3). Six participants were excluded from all physiological analyses: three outliers were identified from inspection of boxplots, electrodermal electrodes became detached during data collection for two participants, and blood pressure was not able to be taken for one person. Physiological data for each measurement window were normally distributed (SWT, p>.05), with the exception of electrodermal activity (AUC) during the story contemplation and moral judgment 1 phase for both congruent and incongruent trials. The t-tests revealed, participants threat-challenge reactivity (TCR) when making Moral judgment 1, t (83) =-2.942, p=.007, 95% CI [-3.433 to -.671], and Moral judgment 2, t (83) =-3.262, p<.005, 95% CI [-3.513 to -.889], was significantly higher in the congruent trials, indicating a greater threat-state (scores <50) when contemplating their choices on the incongruent trials. TCR was not significantly different in the story contemplation phase between congruent and incongruent trials (p>.05). AUC was also significantly higher for incongruent trials in all phases: story contemplation, t (83) = 4.234, p<.001, 95% CI [1.221 to 3.141], moral judgment 1, t (83) = 6.814, p<.001, 95% CI [2.253 to 4.065] and moral judgment 2, t (83) = 8.064, p<.001, 95% CI [2.859 to 4.659], indicating greater autonomic arousal during dilemmas when harm did not maximise outcomes. Despite heart rate being lower for the incongruent trials, this was only significantly

lower for the Moral judgment 2 timeframe i.e. moral acceptability judgments, t (83) = -2.684, p=.012, 95% CI [-2.267 to -.348].

Table 3.

Mean and standard deviation physiological arousal scores for congruent and incongruent trials

		Incongruent trials			Congruent trials		
		Story contemplation	Moral judgment 1	Moral judgment 2	Story contemplation	Moral judgment 1	Moral judgment 2
Threat-Challenge index	*Mean*	49.47	48.99	48.91	50.53	51.01	51.09
	SD	3.74	3.14	3.05	3.74	3.14	3.05
Electrodermal activity (AUC)	*Mean*	51.11	51.59	51.89	48.41	48.41	48.11
	SD	2.40	2.14	2.14	2.402	2.14	2.14
HR	*Mean*	49.68	50.01	49.32	50.32	49.99	50.62
	SD	2.19	2.29	2.21	2.19	2.29	2.27

Note: Scores represent mean and standard deviation *t*-scores for congruent and incongruent trials for story contemplation, moral appropriateness judgment formation (Moral Judgment 1) and moral acceptability judgment formation (Moral judgment 2). NB. 50 is the mean score for *t*-values. Mean scores >50 on incongruent trials indicate greater physiological HR, TCR (challenge state) and AUC relative to congruent trials.

R2.2. Strength of physiological arousal and moral judgments

We tested whether the relative strength of physiological arousal when making moral judgments predicted judgment response tendencies. An arousal-index for the physiological variables for the decision-making measurement windows (Moral judgment 1 and Moral judgment 2) were calculated, by subtracting the mean threat-challenge reactivity (TCR), Area

Under the Curve (AUC), and heart rate (HR) for the congruent trials, from the means on the incongruent trials. Index scores greater than 1 indicate greater mean HR, AUC and TCR on incongruent trials (NB. higher TCR indexes indicate a greater 'challenge' state). Indexes within each measurement window were not correlated. Hierarchical linear regression models were performed for the dependent variables of harm-aversion and outcome-maximisation. The TCR index was inputted at step-1, AUC index at step-2, and HR at step 3, for the moral judgment 1 timeframe. For harm-aversion tendencies, neither Model 1, R^2=.022, F (1, 82) =1.840, p=.179, Model 2, R^2=.026, F (1, 81) =1.068, p=.349, nor Model 3, R^2=.068, F (1, 80) =1.931, p=.131 were significant. For outcome-maximisation tendencies, neither Model 1, R^2=.001, F (1, 82) =.100, p=.752, Model 2, R^2=.025, F (1, 81) =1.039, p=.359, nor Model 3, R^2=.032, F (1, 80) =.886, p=.452, were significant. Therefore, the relative strength of physiological arousal between dilemma types did not predict harm-aversion or outcome-maximisation tendencies on the task.

A further two hierarchical regression models were performed, with acceptability_congruent and acceptability_incongruent scores as the dependent variables. TCR index was inputted at step-1, AUC index at step-2 and HR index at step-3 for the Moral Judgment 2 timeframe. For acceptability_congruent judgments, Model 1 was not significant R^2=.005, F (1, 72) =.387, p=.536. Inclusion of the AUC index in Model 2, R^2=.045, F (1, 71) = 1.666, p=.196, and HR in Model 3, R^2=.048, F (1, 70) = 1.178, p=.324, did not improve model fit. For acceptability_incongruent judgments, Model 1 was not significant R^2=.023, F (1, 72) =1.731, p=.192, but the inclusion of the AUC index in Model 2 led to a significant increase in variance explained, R^2=.075, F (1, 71) =2.897, p=.062, *Sig. F Change*=.05, B=-.042, p=.047, 95% bootstrapped CI [-.082 to -.001]. The addition of HR in model 3 did not significantly improve model fit, R^2=.077, F (1, 70) =1.950, p=.129. Therefore, people exhibiting stronger AUC in the incongruent trials relative to the congruent trials were more likely to rate harm that maximised outcomes as less morally acceptable. With the exception of this finding, the relative strength of physiological arousal people demonstrated between congruent and incongruent trials did not appear to influence their moral judgments.

R3. Interoception and moral judgments of harm

Hierarchical regression analyses were performed to determine whether individual differences in cardiac or gastric interoception influenced moral judgments. Harm-aversion, outcome maximisation, acceptability_congruent and acceptability_incongruent scores, were

the dependent variables. In each regression, the potential confounding variables of age, sex, time-estimation accuracy, state and trait anxiety, and BMI were entered at step-1. Interoceptive accuracy (IAc), Interoceptive sensibility (IS), Interoceptive meta-cognitive awareness (IAc-meta) and gastric interoception were entered at step-2 in separate regression analyses. For harm-aversion, Model 1 was not significant, R^2=.013, F (6, 78) =.178, p=.982. In each regression, the inclusion of IAc, (R^2=.041, F (1, 77) =.466, p=.856), IAc-meta, (R^2=.023, F (1, 77) =.260, p=.967), IS, (R^2=.044, F (1, 77) =.500, p=.832) or gastric interoception, (R^2=.023, F (1, 77) =.260, p=.967) at step-2 did not improve model fit. For outcome-maximisation tendencies, Model 1 was not significant, R^2=.068, F (6, 78) =.947, p=.467. Again, the addition of IAc, (R^2= .068, F (1, 77) = .801, p=.589), IAc-meta, (R^2=.068, F (1, 77) = .801, p=.589), IS, (R^2=.087, F (1, 77) =1.053, p=.402), or gastric interoception (R^2=.070, F (1, 77) =.824, p=.571) at step-2 did not significantly improve model fit. For moral acceptability ratings, Model 1 did not significantly predict acceptability_incongruent judgments, R^2=.103, F (6, 68) =1.303, p=.268. The inclusion of IAc (R^2=.134, F (1, 67) =1.482, p=.189), IAc-meta, (R^2=.127, F (1, 67) =1.398, p=.221), IS, (R^2=.104, F (1, 67) =1.108, p=.368) or gastric interoception, (R^2=.105, F (1, 67) =1.126, p=.358) at step-2 did not significantly improve model fit. For acceptability_congruent judgments, Model 1 was not significant, R^2=.077, F (6, 68) =.945, p=.469. The inclusion of IAc (R^2=.088, F (1, 67) = .925, p=.493), IAc-meta (R^2=.077, F (1, 67) = .801, p=.589), IS (R^2=.105, F (1, 67) = 1.118, p=.363) or gastric interoception (R^2=.078, F (1, 67) =.814, p=.579) did not improve model fit at step-2. Overall, it appears a sensitivity or perceptual awareness of cardiac or gastric sensations did not fundamentally influence moral judgments on the task.

R4. Moderating role of interoceptive accuracy

We explored whether people's level of interoceptive accuracy (IAc) moderated the relationship between the strength of physiological arousal and moral judgments. SPSS scripts for moderation analyses (Model 1; PROCESS) were adopted from Hayes (2018) which applies a bootstrapping method (5000xsamples) as default. Moderation effects are probed at the 16th, 50th and 84th percentiles. The Johnson-Neyman (JN) technique was used to identify regions of significance for moderation (Tabachnick & Fidell, 2013). Interoceptive accuracy (IAc) was entered as the moderator. Each arousal index (TCR, HR, AUC) was entered as a predictor in separate regression models. For clarification, a lower TCR index represents an increased threat-state on incongruent trials (relative to congruent trials).

Harm-aversion and outcome-maximisation

There was a significant interaction between IAc and the TCR index for predicting harm aversion scores, R^2-*chng*=.052, *F* (1, 78) = 4.458, *B*=.0715, *p*=.038 in the model, R^2=.094, *F* (78, 3) =2.696, *p*=.052. The JN technique revealed a lower TCR index (increased threat-state on incongruent trials) predicted reduced harm-aversion tendencies, but only for people scoring above .5698 on the IAc scale, 45.12% of the sample. Therefore, for people who were better at perceiving cardiac sensations, a greater threat-response on incongruent trials (relative to congruent trials) predicted more harm-acceptance responses. When inputting outcome-maximisation as the dependent variable, we did not find an interaction between IAc and TCR index, R^2=.0205, *F* (78, 3) =.5441, *p*=.654. These findings suggest that a certain level of interoceptive accuracy is necessary for cardiovascular threat-reactivity to predict harm-aversion responses. Although higher IAc perceivers did experience a cardiovascular threat response prior to making judgments on incongruent trials, this did not predict an increase in harm-aversive tendencies.

When inputting HR as the predictor for harm-aversion, TCR was entered as a covariate to control for its potential competing influence. The overall model was significant, R^2=.1221, *F* (77, 4) =2.676, *p*=.038. Although the interaction between HR and IAc was not significant, R^2-*chng*=.0345, *F* (1, 77) = 3.023, *B*=. 0917, *p*=.086, the conditional effects of HR on harm-aversion was significant at the 50th [.0017 to .0949] and 84th [.0198 to .1405] percentiles on the IAc scale. The JN technique showed that 51.22% of the sample scoring above .469 on the interoceptive accuracy scale (indicating greater heartbeat counting accuracy), gave more harm-aversive responses if they demonstrated greater anticipatory heart rate on incongruent trials, relative to congruent trials. We also found IAc moderated the relationship between HR and outcome-maximisation tendencies, R^2-*chng*=.051, *F* (1, 78) = 4.274, *B*=-.1096, *p*=.042, in the model, R^2=.0688, *F* (78, 3) =1.921, *p*=.133. This time, a negative relationship between HR and outcome-maximisation tendencies was only significant for people scoring above .796 on the IAc scale, 20.73% of the sample. Therefore, lower anticipatory HR on incongruent trials was associated with greater outcome-maximisation tendencies, but only for people who were particularly good at perceiving cardiac sensations. Finally, when inputting AUC as the predictor of harm-aversion and TCR and HR as covariates, we did not find a significant interaction with IAc in the model R^2=.0963, *F* (76, 5) =1.618, *p*=.165. Similarly, IAc did not moderate the relationship between AUC and outcome maximisation scores (inputting HR as a covariate), in the model, R^2=.0463, *F* (77, 4) =.934,

p=.449. Therefore, the relative strength of anticipatory electrodermal responses did not influence harm-aversion and outcome-maximisation tendencies, regardless of IAc ability. Together, these findings suggest people's ability to detect cardiac sensations may only moderate the relationship between indices of cardiac arousal and moral appropriateness judgments.

Moral acceptability judgments

We next inputted TCR, HR and AUC indexes for the moral judgment 2 window, as predictors in separate regression models. Acceptability_incongruent and acceptability_congruent were the dependent variables. The relationship between TCR and acceptability_congruent judgments was not moderated by interoceptive accuracy, R^2=.0084, *F* (68, 3) =.1914, *p*=.902. There was no interaction between IAc and TCR for acceptability_incongruent judgments, R^2-*chng*=.015, *F* (1, 68) = 1.163, *p*=.285. However, the coefficient for IAc alone was significant (*B*=.474, *p*=.012) in the model, R^2=.111, *F* (68, 3) =2.822, *p*=.045, suggesting IAc may influence acceptability judgments on congruent trials, but only when cardiovascular threat-reactivity scores remain constant. There was no interaction between HR and IAc for acceptability_incongruent, R^2=.093, *F* (68, 3) =2.325, *p*=.082, or acceptability_congruent scores, R^2=.031, *F* (68, 3) =.732, *p*=.537. The relationship between AUC and acceptability_congruent scores was also not moderated by IAc, R^2=.0732, *F* (68, 3) =1.789, *p*=.157. However, we did find an interaction between IAc and AUC for predicting acceptability_incongruent scores, R^2-*chng*=.0461, *F* (1, 68) = 3.871, *B*=.086, *p*=.053, in the model R^2=.189, *F* (68, 3) =5.314, *p*=<.005. The JN technique revealed higher AUC on incongruent trials was associated with greater disapproval of harmful acts that maximised outcomes, but only for people scoring below .440 on the IAc scale, 47.22% of the sample. Therefore, autonomic arousal was associated with greater moral disapproval of harmful acts that maximised outcomes, but only for people with relatively lower levels of interoceptive accuracy. To rule out any confounding influence of trait anxiety on IAc, we included trait anxiety as a covariate in all models showing significant IAc moderation. Trait anxiety did not significantly change any of the moderation effects found and provided no significant additional explanatory power.

Exploratory analyses

ER1: Does physiological arousal predict response-time on incongruent trials, and is this moderated by interoception?

We tested whether relative physiological arousal predicted the difference in response time for congruent and incongruent trials. To create response time indexes, we subtracted mean response time (RT) for the congruent trials from the mean RT for the incongruent trials for the Moral judgment 1 (appropriateness) and Moral judgment 2 (acceptability) timeframes. Lower scores indicate longer time-taken to make judgments on the incongruent trials. We also explored moderation effects of interoceptive ability. As we had no clear hypotheses about which dimensions of interoception would be relevant, IAc, IS, IAc-meta and gastric interoception were inputted as moderators in separate regression analyses using PROCESS Model 1 (Hayes, 2018).

Moral appropriateness judgments

We inputted the arousal indexes as predictors of RT for moral judgment 1 into a bootstrapped (x1000) hierarchical linear regression model; TCR index (step-1), AUC index (step-2) and HR index (step 3). Casewise diagnostics identified two outliers that were removed. In Model 1, TCR did not significantly predict RT for moral judgment 1, R^2=.001*,* F (80, 1) =.073, p=.788, but the addition of the AUC index in Model 2 led to a significant increase of variance explained, R^2=.119*,* F (79, 1) =5.337, p=.007. The addition of HR in model 3 did not significantly improve model fit, Sig *F Chn*g=.447, but the model was significant R^2=.126*,* F (78, 1) =3.734, p=.015. AUC represents the strength of electrodermal activity between question presentation and response, divided by response time. Therefore, a stronger and more sustained electrodermal response when making moral appropriateness judgments on incongruent trials, was associated with longer response times for these trials. We did not find any of the interoception variables moderated the relationship between AUC and RT, (F<1, p>.05).

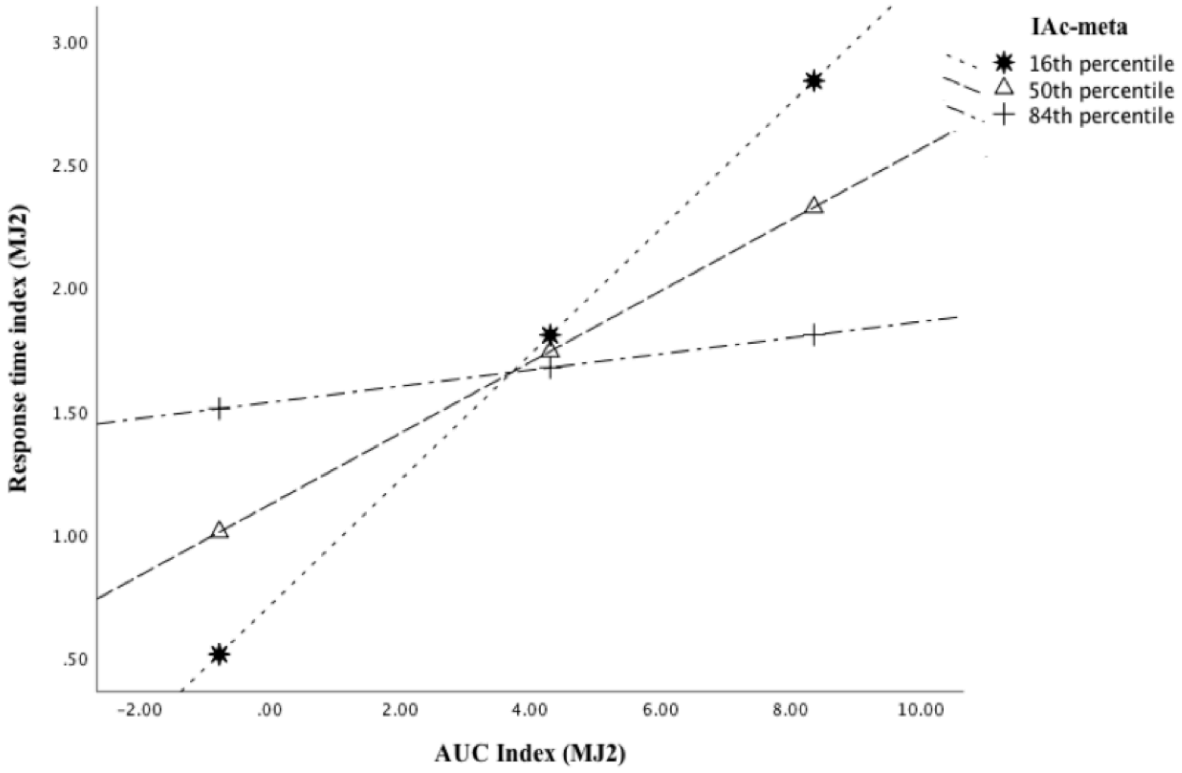
Moral acceptability judgments

We repeated the former analysis for RT in the moral judgment 2 timeframe, inputting the arousal indexes into the hierarchical regression as before. Casewise diagnostics revealed 3 outliers that were removed. Model 1 including just the TCR index did not predict RT for

moral judgment 2, R^2=.009, F (79, 1) =.716, p=.400. As before, the inclusion of AUC in Model 2, led to a significant increase in variance explained, R^2=.158, F (78, 1) =7.329, p=.001. Inputting HR in Model 3 did not significantly improve model fit, Sig *F Chng*= .582, but the model was significant, R^2=.162, F (77, 1) =4.944, p=.003. Again, a more sustained electrodermal response when making moral acceptability judgments on incongruent trials (where harm maximises outcomes) predicted longer response times on these trials. Interestingly, here we found both IAc-meta, R^2 *chng*=.0512, F (75, 1) =4.837, p=.031 in the model, R^2=.206, F (75, 3) =6.499, p=.0006, and gastric interoception R^2*chng*=.0432, F (75, 1) =4.073, p=.047 in the model, R^2=.204, F (75, 3) =6.398, p=.0006, moderated the relationship between AUC and response time. A reduced awareness of one's ability to accurately perceive cardiac sensations (IAc-meta) strengthened the relationship between AUC and response time i.e. stronger electrodermal responses predicted even longer reaction times (Figure 1). The JN technique revealed moderation was present for people scoring below .313 on the IAc-meta scale, 69.62% of the sample; indicating this effect did not apply to people very aware of their interoceptive accuracy. For gastric interoception, stronger electrodermal responses were also associated with longer response times for people with a reduced sensitivity to stomach sensations. However, the JN technique showed moderation was not significant for people with very high gastric sensitivity i.e. those scoring below 31.83 on the gastric interoception scale, 21.52% of the sample (Figure 2).

Figure 1.

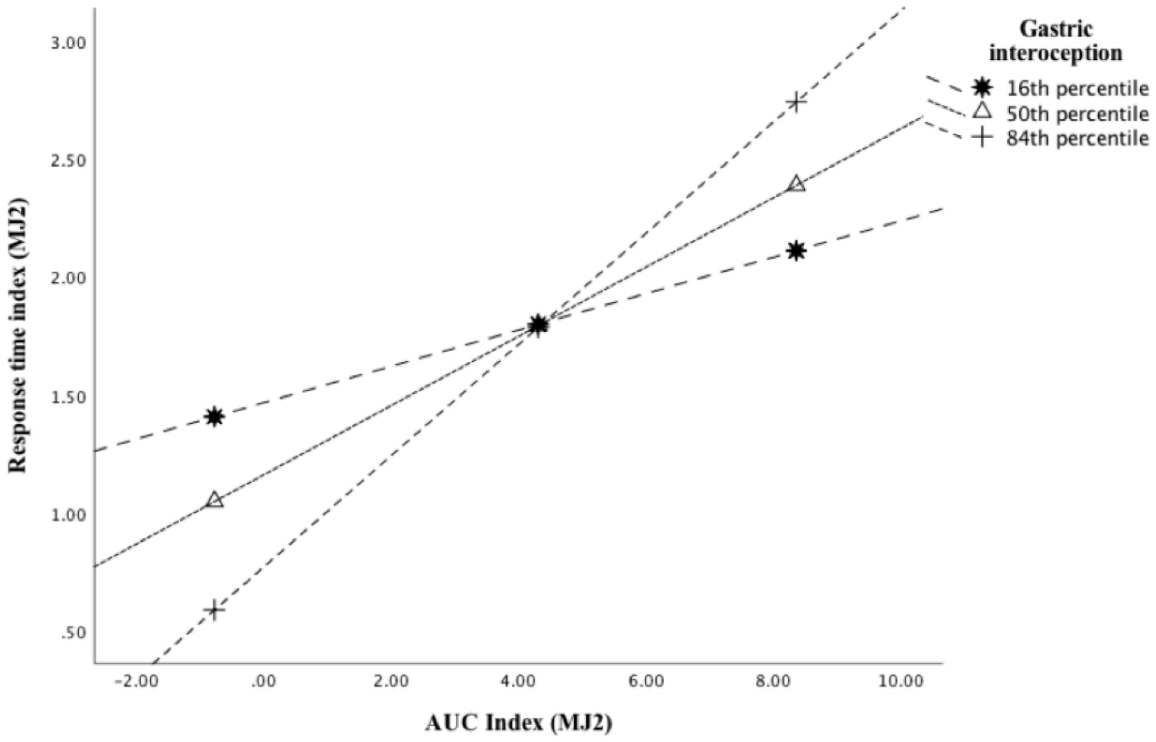
Regression lines depicting moderation effect of interoceptive meta-cognitive awareness (IAc-meta) in relationship between AUC index and Response time (RT) index (RT incongruent - RT congruent) for moral acceptability judgments (MJ2)



Note: Legend symbols represent moderation effects probed at 16th, 50th and 84th percentiles of IAc-meta. Moderation was significant (p<.05) at 16th and 50th percentiles.

Figure 2.

Regression lines depicting moderation effect of gastric interoception in relationship between AUC index and Response time (RT) index (RT incongruent - RT congruent) for moral acceptability judgments (MJ2)



Note: Legend symbols represent moderation effects probed at 16th, 50th and 84th percentiles of gastric interoception. Moderation was significant (p<.05) at 50th and 84th percentiles.

Discussion

Interoceptive processes are central to shaping motivation, emotional experience and facilitating adaptive behavioural responses to our environment (Craig, 2015; Feldman Barrett, 2017; Friston, 2010). Despite the fundamental influence of emotional and physiological processes in moral decision-making (Cushman et al., 2012; Damasio et al., 1990; Damasio, 1996; Greene et al., 2001; Moretto et al., 2010; Parton & McGinley, 2019), relatively little is known about how individual differences in interoceptive experiences could influence the relationship between physiological processes and moral judgments. This is the first study to examine the role of interoception in the relationship between anticipatory physiological arousal and moral judgments of harm. Our results suggest interoceptive accuracy in

particular, plays a role in moderating the effect of physiological arousal on moral judgments of harm. Interestingly, we found the moderated effects of heart rate, cardiovascular threat-reactivity and electrodermal responses showed diverging relationships with moral judgments. This suggests that discrete somatic markers may be associated with distinct psychological processes leading up to moral judgments, which can be enhanced or attenuated by an ability to accurately perceive visceral sensations.

Contrary to some prior research (Herbert et al., 2012; Whitehead & Drescher, 1980), we did not find any association between interoceptive accuracy (IAc) and any measures of gastric interoception (R1), indicating a sensitivity to visceral sensations within the cardiac and gastrointestinal systems were distinct (Ferentzi et al., 2018a). Although, this predictive relationship was just below significance (p=.063). Further higher-powered studies are necessary to determine whether individual differences in interoceptive sensitivity are consistent across visceral systems. None of the interoception variables predicted any of the moral judgment variables (R3), suggesting a perceptual awareness or sensitivity to internal sensations did not fundamentally influence moral judgments, contrary to previous work showing a link between gastric-related sensitivity (Schnall et al., 2008; Vicario et al., 2018), and IS (Brown et al., 2020) with moral judgments. We did find that the physiological arousal parameters were higher in incongruent compared to the congruent trials (R2.1). Specifically, electrodermal responses and sympathetic cardiovascular arousal was significantly higher when making appropriateness and acceptability judgments on the incongruent trials compared to the congruent trials. Participants were more likely to condone harm on the incongruent trials, which suggests anticipatory physiological reactions were higher prior to harm-acceptance judgments. Heart rate was lower for incongruent trials, but only significantly lower when contemplating moral acceptability judgments. This is consistent with previous work finding an association between heart rate deceleration when contemplating dilemmas of direct interpersonal harm (Carmona-Perera et al., 2013), and suggests participants were experiencing a stronger negative emotional reaction towards incongruent dilemmas (Lang et al., 1993).

Despite these differences in physiological arousal, the relative difference in physiological arousal between incongruent and congruent trials, was not associated with people's harm aversion and outcome maximisation tendencies (R2.2). However, in R4, we found several moderation effects of interoceptive accuracy (IAc) in the relationship between

arousal and moral judgments. On incongruent trials: increased cardiovascular threat-reactivity predicted reduced harm-aversion tendencies but only for higher IAc perceivers; increased HR predicted increased harm-aversion tendencies and decreased HR predicted increased outcome-maximisation tendencies but only for higher IAc perceivers. Finally, increased electrodermal arousal predicted stronger disapproval of harmful acts that maximised outcomes but only for relatively lower IAc perceivers. These findings are discussed further below.

Strength of physiological arousal and moral judgments

We did not find the relative strength of physiological arousal on congruent versus incongruent trials predicted moral judgments, with the exception of electrodermal responses which predicted lower moral acceptability ratings for incongruent trials. There is some evidence to suggest the strength of physiological arousal may not always predict moral judgments of harm (Cecchetto et al., 2018; Francis et al., 2018), but other evidence suggests the strength or frequency of electrodermal responses (e.g. McDonald et al., 2017; Moretto et al., 2010) and individual differences in cardiovascular threat-reactivity (Cushman et al., 2012) predicts harm rejection responses in personal moral dilemma scenarios. However, prior psychophysiological studies did not use a 'congruent' comparison dilemma to calculate the difference in physiological arousal for 'incongruent' trials. Furthermore, the moral dilemma stimuli in the current study (Conway & Gawronski, 2013) included a majority of situations involving personal harm, and several situations that involved more indirect harm. As negative emotional and physiological arousal has previously been associated with an aversion to the former (Carmona-Perera et al., 2013; Cushman et al., 2012; Greene et al., 2001; McDonald et al., 2017), it is possible that the dilemmas involving direct interpersonal harm were responsible for the significantly higher physiological arousal for the incongruent trials. If arousal was not significantly higher on incongruent trials involving more indirect harm (compared to the matched congruent trials), this would explain the lack of correlation between our arousal measures and judgment response tendencies.

Moderating effect of IAc on relationship between arousal and moral judgments

Although the relative difference in physiological arousal on incongruent trials did not predict outcome-maximisation or harm-aversion tendencies, we found these relationships were moderated by people's ability to perceive cardiac sensations (IAc). Greater

cardiovascular threat-reactivity on incongruent trials was associated with reduced harm-aversion tendencies, but only for people with relatively higher interoceptive accuracy, ~45% of the sample. This suggests that a greater probability of accepting harmful actions overall, was associated with a greater threat response on incongruent trials, and likely represents an anticipatory aversive response to the prospect of accepting harm. In other words, greater threat reactivity did not deter these people from accepting harm but instead appears to be associated with the prospect of accepting harm. We found a similar moderation effect with heart rate (HR), whereby increased HR on incongruent trials was associated with an increase in harm-aversive response tendencies, but only for ~50% of people with higher IAc ability. This is consistent with the moderation effect for cardiovascular threat-reactivity, as increases in heart rate are associated with increases in cardiac output, which was used to calculate the cardiovascular threat-reactivity index.

These findings suggest that cardiovascular threat reactivity may be associated with harm-aversion response tendencies towards both 'personal' harmful acts (Cushman et al., 2012) and more impersonal harmful acts, at least for higher IAc perceivers. Conversely, higher IAc perceivers demonstrating relatively higher cardiac output to systemic vascular resistance (associated with cardiovascular 'challenge' states) on incongruent trials was associated with an increase in harm-aversion response tendencies. It is possible that better heartbeat detectors appraised the incongruent dilemmas as more of a 'challenge' than the congruent dilemmas, as these dilemmas involved a moral conflict that was absent from the congruent dilemmas. However, on average people demonstrated significantly greater cardiovascular threat-reactivity in the pre-decision timeframe on incongruent trials, compared to congruent trials, suggesting these trials were generally experienced as more physiologically aversive.

The finding that a greater difference in cardiovascular reactivity on incongruent trials, was only associated with overt harm-aversion tendencies for people with relatively better heartbeat detection ability, is consistent with evidence showing heartbeat detection ability predicts heightened reactivity to emotional stimuli (Eichler & Katkin, Edward, 1994; Pollatos et al., 2007b). Ainley et al (2016) suggest that because individuals who are better at detecting heartbeats update interoceptive priors more often (due to more precise prediction errors associated with heartbeat detection ability), they will demonstrate greater autonomic reactivity to an affective stimulus if the stimulus influences their heartbeat in a way that is not completely predicted, and subsequently results in interoceptive prediction errors. Greater

changes in cardiovascular arousal on incongruent trials is likely to evoke larger prediction errors, which may explain why better heartbeat detectors were more autonomically reactive to the moral dilemma stimuli, which in turn meant their overt moral judgments were associated with cardiovascular arousal (Ainley et al., 2016). However, higher IAc individuals appear to have been less influenced by increases in aversive psychophysiological arousal (threat reactivity and lower heart rate) on incongruent trials, which has shown to lead people to reject harmful action in traditional moral dilemmas of personal harm, such as the Footbridge dilemma (Carmona-Perera et al., 2013; Cushman et al., 2012; Greene et al., 2001; Moretto et al., 2010). If IAc was associated with enhanced accuracy of interoceptive representations associated with perceptual forms of inference (Farb et al., 2015; Seth & Friston, 2016), this may have tempered regulatory behaviours aimed at alleviating a negative arousal state (i.e. rejecting harm). Therefore, better heartbeat detectors may demonstrate superior interoceptive regulatory processes, that allow them to accept harmful actions even if it feels physiologically aversive to do so. Consistent with this, better heartbeat detectors have shown to be more effective at regulating negative emotional states (Füstös et al., 2013) and IAc predicts both antecedent and response-focused emotional regulation strategies (Kever et al., 2015). Action aversion (aversion to harmful actions) and outcome aversion (aversion to negative consequences of harm) have both been associated with harm-aversion tendencies using this moral dilemma task (Miller et al., 2014; Reynolds & Conway, 2018). However, cardiovascular threat reactivity (systemic vascular resistance) has previously been associated with an aversive response to performing harmful actions and not with witnessing the consequences of harm (Cushman et al., 2012), therefore IAc may be uniquely important in regulating the influence of threat-reactivity associated with action-aversion, on harm-aversion response tendencies.

IAc did not influence the relationship between cardiac threat-reactivity and outcome-maximisation tendencies and is consistent with prior work showing vagally-mediated measures of cardiac reactivity (i.e. resting heart rate variability) are associated with outcome-maximisation tendencies but not harm-aversion tendencies (Park et al., 2016). We did find that lower HR on incongruent trials, predicted greater outcome-maximisation tendencies for ~20% of the sample scoring particularly high on IAc. A greater negative emotional response, potentially characterised by heart rate deceleration (Lang et al., 1993), is believed to occur during the evaluation of moral violations (Rozin et al, 1999), particularly in dilemmas involving personal harm (Carmona-Perera et al., 2013). Painful injury-related disgust has also

been associated with heart rate deceleration (Shenhav & Mendes, 2014). Changes in cardiac sensations across congruent and incongruent trials may have been more accessible for higher IAc perceivers, strengthening the relationship between cardiovascular arousal and outcome-maximisation tendencies. There is evidence to suggest that deficits in the perception of somatic markers (Carmona-Perera et al., 2013), results in an absence of cardiac reactivity (HR) when contemplating dilemmas of personal harm, resulting in more utilitarian judgments in traditional moral dilemmas. However, whether an increase in utilitarian judgments was related to a reduced aversion to harm or increased motivation to maximise outcomes was unclear (Carmona-Perera et al., 2013). Our findings suggest that people with a superior ability to perceive cardiac sensations, may more effectively modulate their heart rate when contemplating moral dilemmas that involve personal and impersonal harm, which predicts an increase in outcome-maximisation tendencies and a reduction in harm-aversion tendencies.

We did not find IAc moderated the relationship between electrodermal responses and harm-aversion or outcome-maximisation tendencies. However, electrodermal responses did predict lower moral acceptability ratings on incongruent trials for people scoring relatively lower on IAc, 47.22% of the sample. Therefore, greater autonomic arousal on incongruent trials was associated with stronger disapproval of harmful acts that maximised outcomes, but only for people who were relatively poorer at perceiving internal sensations. This is counterintuitive to the work of Dunn et al (2010) who found that IAc strengthened the relationship between somatic markers and intuitive decision-making. This is potentially because the moral dilemma task does not provide the opportunity for people to learn about the reliability of somatic markers for predicting positive or negative outcomes, which is fundamentally different to the intuitive reasoning task used by Dunn et al (2010). In the context of moral dilemmas, one possibility is that people with less accurate interoceptive representations (lower IAc) may engage in forms of active inference when faced with surprising changes in autonomic arousal (Farb et al., 2015). Specifically, stronger disapproval of violations of harm may have acted as a means of sensory regulation, to alleviate an aversive emotional response associated with harmful action (Farb et al., 2015). Overall these moderation effects suggest that a greater ability to perceive internal sensations can sometimes strengthen and sometimes weaken the relationship between physiological arousal and moral judgments. IAc may be particularly relevant for selectively regulating the influence of somatic signals (Füstös et al., 2013; Kever et al., 2015) on moral judgments, depending on the nature or utility of the signal at the time. Further work is needed to establish the

robustness of these effects, and the potential adaptive influence of IAc in the relationship between arousal and moral cognition.

Arousal, interoception and response time

An exploratory analysis examined whether the link between physiological arousal and response time was moderated by interoception. We found that a stronger, more sustained electrodermal response on incongruent trials predicted longer reaction times for both moral appropriateness and moral acceptability judgments. Greene et al's (2001) original fMRI study found longer reaction times preceded harm acceptance judgments in personal moral dilemmas that they attributed to the 'emotional incongruence' of their response, i.e. condoning immoral action was counterintuitive to their emotional response. In the current study, greater sympathetic autonomic arousal generated on incongruent trials, may have predicted longer reaction times if people were more conflicted about choosing a 'utilitarian' response that was incongruent with their strong emotional reaction to the prospect of harm. The link between electrodermal response and moral appropriateness judgments was not moderated by interoception. However, we found stronger electrodermal responses were associated with even longer reaction times for moral acceptability judgments for people lower in gastric interoception and lower in IAc-meta (an awareness of one's interoceptive accuracy ability). Importantly, these moderation effects were not significant for people very low in gastric sensitivity and very high in IAc-meta.

Our measure of gastric interoception captured how early people noticed gastrointestinal processes associated with stomach-filling. People with lower scores were aware of sensations of gastric distension much sooner in the drinking process, before reaching maximum fullness. Gastrointestinal processes such as hunger and disgust sensitivity (Schnall et al., 2008; Vicario et al., 2018) have shown to influence allocentric judgments of ethical violations in previous studies, potentially due to the link with disgust-related emotions (Tracy et al., 2019). Gastric myoelectrical activity strongly correlates with measures of arousal following exposure to emotional stimuli and appears to be associated with changes in the sympathetic nervous system (Vianna & Tranel, 2006). Electrodermal responses have also been associated with disgust (Rohrmann & Hopp, 2008). These associations may explain the interaction between gastric interoception and electrodermal responses. If a sensitivity to stomach sensations heightened people's awareness to disgust sensations, which was reciprocated by an increase in electrodermal-activity, this could reduce reaction time i.e. if

disgust-related somatic signals were more accessible as a form of intuitive guidance, this might precipitate a swifter, more confident response. As we did not measure myoelectrical gastric processes, it is unclear how gastric interoception may have been influenced by concurrent gastrointestinal signals.

The IAc-meta moderation effect is also consistent with some recent work by Vega and colleagues who found meta-cognition (expressed as a confidence rating) influenced how 'right' people felt their moral decisions to be, which influenced choices to rethink their answer and overall response time (Vega et al., 2020). A reduced awareness of one's ability to accurately perceive visceral sensations could indicate lower confidence in the reliability of somatic markers associated with task-related arousal; resulting in longer deliberation times to determine a judgment that 'felt' right. IAc meta may therefore contribute to people's sense of somatic intuition about the 'rightness' (Thompson, 2009; Vega et al., 2020) of their moral choices. Together, these novel findings suggest gastric interoception and IAc-meta may influence how quickly somatic signals associated with autonomic arousal translate into action, when making allocentric judgments about harmful acts that maximise outcomes.

Limitations

We acknowledge several limitations of this study. As we were interested in exploring the broader influence of interoceptive and physiological processes on people's harm-based response tendencies (Conway & Gawronski, 2013), we used a moral judgment measure that incorporates several situations of impersonal harm. As prior research has shown a stronger link between the physiological parameters and dilemmas involving personal harm, this may be why we did not find linear associations between the strength of physiological arousal and moral judgments. In addition, we did not measure myoelectrical gastric processes during the moral judgment task which prevents us from exploring interactions between gastric somatic signals and gastric interoception during the moral judgment task. In addition, future iterations would benefit from trait measures of action-aversion and outcome-aversion (Miller et al., 2014), to better understand how these motivational factors may interact with IAc, physiological arousal and moral judgments. Furthermore, the validity of the heartbeat detection task (Schandry, 1981) has been challenged by evidence that beliefs about heart rate can influence performance on this task (Kleckner et al., 2015; Ring et al., 2015; Ring & Brener, 2018). In addition, future replications of this study could include an attention check to rule out the possibility that engagement with the task influenced moral judgments in any

way. Lastly, although we had a reasonable sample size for psychophysiological studies, further work is needed to demonstrate the robustness of these novel findings in different populations. Exploring the relationship between physiological arousal and interoception in more immersive, emotive virtual-reality moral dilemmas could shed light on the role of bodily perception in moral behaviour in more ecologically valid settings.

Conclusion

This study explored the role of interoceptive processes in the link between physiological arousal and moral judgments of harm. Fundamentally we have shown how individual differences in the perception of visceral sensations, can shape how physiological signals influence moral decision-making. Cardiac and gastric interoception did not fundamentally influence moral judgments, however, greater accuracy in perceiving cardiac sensations (IAc) moderated the link between cardiovascular arousal and egocentric moral judgments of harm, and between electrodermal-activity and allocentric moral judgments of harm. Interestingly, the moderation effects of IAc were different for each physiological parameter, suggesting IAc may be important in selectively enhancing or diminishing the effect of physiological arousal on decision-making, depending on the nature of the somatic signal. We suggest IAc may indicate a reduced reactivity towards aversive physiological arousal that allows people to accept harmful actions, in spite of an aversive state. IAc may also facilitate more adaptive modulation of heart rate when presented with aversive moral dilemma stimuli, which leads to an increase in outcome-maximisation response tendencies. The moderating role of gastric interoception and interoceptive meta cognitive awareness in the relationship between electrodermal activity and response time is intriguing and warrants further investigation. Importantly, we did not find that interoception was influential for moral judgments alone, and the moderation effects found were conditional upon a certain level of interoceptive accuracy. Therefore, individual differences across other measures of interoception, such as sensibility and awareness, may be less relevant to moral decision-making processes. This study provides many avenues for future research to explore the role of interoception in moral cognition, potentially investigating its broader role in anticipatory emotional processes present in other types of moral decision-making.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

H.B. conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper.

Acknowledgments

We thank Dr Mary Nikolaidou (University of Bath) for technical support with BIOPAC setup. We thank Dr Susanna Martin (University of Bath) for lab support and Caroline Ransford (University of Bath) for support with recruitment. We thank Dr Harry Farmer (University of Greenwich) for useful discussions. We thank Steven Clifford (BIOPAC) for support with processing and scripting for physiological data analysis. This studentship runs alongside UK EPSRC grant Virtual Realities: Immersive Documentary Encounters (ref: EP/P025595/1).

References

Ainley, V., Apps, M. A. J., Fotopoulou, A., & Tsakiris, M. (2016). “Bodily precision”: A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0003

Ainley, V., & Tsakiris, M. (2013). Body Conscious? Interoceptive Awareness, Measured by Heartbeat Perception, Is Negatively Correlated with Self-Objectification. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0055568

Anderson, E. R., & Hope, D. A. (2009). The relationship among social phobia, objective and

perceived physiological reactivity, and anxiety sensitivity in an adolescent population. *Journal of Anxiety Disorders*, *23*(1), 18–26. https://doi.org/10.1016/j.janxdis.2008.03.011

Anderson, S. W., Bechara, A., Damásio, H., Tranel, D., & Damásio, A. R. (2013). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Social Neuroscience: Key Readings*, *9780203496*(11), 29–40. https://doi.org/10.4324/9780203496190

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801. https://doi.org/10.1037/xge0000533

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2014.09.003

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, *275*(5304), 1293 LP – 1295. http://science.sciencemag.org/content/275/5304/1293.abstract

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage*, *31*(2), 945–950. https://doi.org/10.1016/j.neuroimage.2005.12.039

BIOPAC Systems, I. (2015). *Acknowledge Demo Guide*.

BIOPAC Systems, I. (2020). *MP160 BIOPAC Sytems*. https://www.biopac.com/product/mp150-system-221/

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *J.P. Forgas (Ed.) Feeling and thinking: The role of affect in social cognition* (Ed., pp. 59–82). Cambridge University Press.

Blascovich, J., & Mendes, W. B. (2010). Social Psychophysiology and Embodiment. *Handbook of Social Psychology*, 194–227. https://doi.org/10.1002/9780470561119.socpsy001006

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Braithwaite, J. J., & Watson, D. G. (2015). Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs). *Technical Research Note.* Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham ©

Braithwaite, J. J., Watson, D. G., Robert, J., & Mickey, R. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. …. *Technical Report.* Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham © https://doi.org/10.1017.S0142716405050034

Brown, H., Proulx, M. J., & Fraser, D. S. (2019). *Does interoception predict moral judgments of harm?* Open Science Framework. https://osf.io/m6f5j

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). Hunger Bias or Gut Instinct ? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making. *Frontiers in Psychology: Perception Science*, *11*(September), 1–16. https://doi.org/10.3389/fpsyg.2020.02261

Carmona-Perera, M., Reyes del Paso, G. A., Pérez-García, M., & Verdejo-García, A. (2013). Heart rate correlates of utilitarian moral decision-making in alcoholism. *Drug and Alcohol Dependence*, *133*(2), 413–419. https://doi.org/10.1016/j.drugalcdep.2013.06.023

Cecchetto, C., Korb, S., Rumiati, R. I., & Aiello, M. (2018). Emotional reactions in moral decision-making are influenced by empathy and alexithymia. *Social Neuroscience*, *13*(2), 226–240. https://doi.org/10.1080/17470919.2017.1288656

Clark, D. M., Salkovskis, P. M., Breitholtz, E., Westling, B. E., Öst, L. G., Koehler, K. A., Jeavons, A., & Gelder, M. (1997). Misinterpretation of body sensations in panic

disorder. *Journal of Consulting and Clinical Psychology*, *65*(2), 203–213. https://doi.org/10.1037/0022-006X.65.2.203

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235. https://doi.org/10.1037/a0031021

Craig, B. A. D. (2015). *How do you Feel?:An Interoceptive Moment with Your Neurobiological Self*. Princeton University Press.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*(2), 189–195. https://doi.org/10.1038/nn1176

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: the aversion to harmful action. *Emotion (Washington, D.C.)*, *12*(1), 2–7. https://doi.org/10.1037/a0025071

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81–94.

Damasio, A.R. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

De Berardis, D., Campanella, D., Gambi, F., La Rovere, R., Sepede, G., Core, L., Canfora, G., Santilli, E., Valchera, A., Mancini, E., Salerno, R. M., Moschetta, F. S., & Ferro, F. M. (2007). Alexithymia, fear of bodily sensations, and somatosensory amplification in young outpatients with panic disorder. *Psychosomatics*, *48*(3), 239–246. https://doi.org/10.1176/appi.psy.48.3.239

Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, *30*(1), 1–11. https://doi.org/10.1016/j.cpr.2009.08.008

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., & Dalgleish, T. (2010). Listening to Your Heart: How Interoception Shapes Emotion Experience and Intuitive Decision Making. *Psychological Science*, *21*(12), 1835–1844. https://doi.org/10.1177/0956797610389191

Duschek, S., Werner, N. S., Reyes del Paso, G. A., & Schandry, R. (2015). The Contributions of Interoceptive Awareness to Cognitive and Affective Facets of Body Experience. *Journal of Individual Differences*, *36*(2), 110–118. https://doi.org/10.1027/1614-0001/a000165

Eichler, S., Katkin, Edward, S., Blascovich, J., & Kelsey, R. M. (1987). Cardiodynamic factors in heartbeat detection and the experience of emotion [Abstract]. *Psychophysiology*, *24*, 587.

Eichler, S., & Katkin, E. S. (1994). The relationship between cardiovascular reactivity and heartbeat detection. *Psychophysiology*, *31*(3), 229–234. https://doi.org/10.1111/j.1469-8986.1994.tb02211.x

Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., & Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in Psychology*, *6*(JUN), 1–26. https://doi.org/10.3389/fpsyg.2015.00763

Feldman Barrett, L. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsw154

Feldman Barrett, L., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional. *J Pers Soc Psychol.*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Feldman Barrett, L., & Simmons, W. K. (2015). Interoceptive predictions in the brain. In *Nature Reviews Neuroscience* (Vol. 16, Issue 7). https://doi.org/10.1038/nrn3950

Ferentzi, E., Bogdány, T., Szabolcs, Z., Csala, B., Horváth, Á., & Köteles, F. (2018a). Multichannel investigation of interoception: sensitivity is not a generalizable feature. *Frontiers in human neuroscience*, 12, 223.

Ferentzi, E., Drew, R., Tihanyi, B. T., & Köteles, F. (2018b). Interoceptive accuracy and body awareness – Temporal and longitudinal associations in a non-clinical sample. *Physiology and Behavior*, *184*(August 2017), 100–107.

https://doi.org/10.1016/j.physbeh.2017.11.015

Foot, P. (2003). The Problem of Abortion and the Doctrine of the Double Effect. *Virtues and Vices*, *5*, 19–32. https://doi.org/10.1093/0199252866.003.0002

Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., & Schulz, A. (2016). Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology*, *109*(November), 71–80. https://doi.org/10.1016/j.ijpsycho.2016.09.019

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2018). Virtual morality in the helping professions: Simulated action and resilience. *British Journal of Psychology*, *109*(3), 442–465. https://doi.org/10.1111/bjop.12276

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Anderson, G. G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0164374

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. https://doi.org/10.1257/089533005775196732

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: Interoceptive awareness facilitates reappraisal. *Social Cognitive and Affective Neuroscience*, *8*(8), 911–917. https://doi.org/10.1093/scan/nss089

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: “Anterior insular cortex mediates bodily sensibility and social anxiety” by Terasawa et al.,(2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234. https://doi.org/10.1093/scan/nss140

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 65–74. https://doi.org/10.1016/j.biopsycho.2014.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105 LP – 2108. http://science.sciencemag.org/content/293/5537/2105.abstract

Gu, J., Zhong, C. B., & Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *Journal of Experimental Psychology: General*, *142*(2), 307–312. https://doi.org/10.1037/a0029549

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2013). Emotional regulation and bodily sensation: Interoceptive awareness is intact in borderline personality disorder. *Journal of Personality Disorders*, *27*(4), 506–518. https://doi.org/10.1521/pedi_2012_26_049

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd ed.). The Guildford Press: New York.

Herbert, B. M., Muth, E. R., Pollatos, O., & Herbert, C. (2012). Interoception across modalities: On the relationship between cardiac awareness and the sensitivity for gastric functions. *PLoS ONE*, *7*(5), 1–9. https://doi.org/10.1371/journal.pone.0036646

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Kahane, G. (2014). Intuitive and Counterintuitive Morality. *Moral Psychology and Human Agency*, 9–39. https://doi.org/10.1093/acprof:oso/9780198717812.003.0002

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kant, I. (2018). Groundwork of the Metaphysics of Morals (1785). *Business Ethics: The Big*

Picture, 18.

Kever, A., Pollatos, O., Vermeulen, N., & Grynberg, D. (2015). Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. *Personality and Individual Differences*, *87*. https://doi.org/10.1016/j.paid.2015.07.014

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology*, *52*(11), 1432–1440. https://doi.org/10.1111/psyp.12503

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.4324/9781315853161

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at Pictures: Affective, facial, visceral and behavioural reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1080/0449010X.1992.10705887

McDonald, M. M., Defever, A. M., & Navarrete, C. D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior*, *38*(6), 770–778. https://doi.org/10.1016/j.evolhumbehav.2017.06.001

Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PLoS ONE*, *13*(12), 1–12. https://doi.org/10.1371/journal.pone.0208034

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the Unexpected: Physiological Responses During Social Interactions With Expectancy-Violating Partners. *Journal of Personality and Social Psychology*, *92*(4), 698–716. https://doi.org/10.1037/0022-3514.92.4.698

Mill, J. S. (1998). *Utilitarianism* (R. Crisp,). Oxford University Press.

Miller, L.C., Murphy, R. and Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology*, *41*(2), 397.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion (Washington, D.C.)*, *14*(3), 573—587. https://doi.org/10.1037/a0035361

Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–1899. https://doi.org/10.1162/jocn.2009.21367

Naqvi, N. H., & Bechara, A. (2006). Skin conductance: A psychophysiological approach to the study of decision making. In C. Senior, T. Russell & M. Gassaniga (Ed.), *Methods in mind* (pp. 103–122). https://doi.org/10.1088/2058-7058/26/05/35

Nentjes, L., Meijer, E., Bernstein, D., Arntz, A., & Medendorp, W. (2013). Brief Communication: Investigating the Relationship Between Psychopathy and Interoceptive Awareness. *Journal of Personality Disorders*, *27*(5), 617–624. https://doi.org/10.1521/pedi_2013_27_105

OMRON, I. (n.d.). *OMRON M3 Comfort*. www.omron-healthcare.co.uk/blood-pressure-monitors/m3_comfort_2.html

Palomba, D., Sarlo, M., Angrilli, A., Mini, A., & Stegagno, L. (2000). Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology*, *36*(1), 45–57. https://doi.org/10.1016/S0167-8760(99)00099-9

Park, G., Kappes, A., Rho, Y., & Van Bavel, J. J. (2016). At the heart of morality lies neuro-visceral integration: Lower cardiac vagal tone predicts utilitarian moral judgment. *Social Cognitive and Affective Neuroscience*, *11*(10), 1588–1596. https://doi.org/10.1093/scan/nsw077

Parton, D. M., & McGinley, J. J. (2019). At the heart of harm: Cardiac substrates of action-based aversion to harm. *Personality and Individual Differences*, *151*(July 2018), 109497. https://doi.org/10.1016/j.paid.2019.07.007

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2020). Reasoning Supports Utilitarian Resolutions to Moral Dilemmas Across Diverse Measures. *Journal of Personality and Social Psychology*, 1–42. https://doi.org/10.1037/pspp0000281

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure & Function*, *214*(5–6), 451–463. https://doi.org/10.1007/s00429-010-0258-

Pollatos, O., Herbert, B. M., Kaufmann, C., Auer, D. P., & Schandry, R. (2007a). Interoceptive awareness, anxiety and cardiovascular reactivity to isometric exercise. *International Journal of Psychophysiology*, *65*(2), 167–173. https://doi.org/10.1016/j.ijpsycho.2007.03.005

Pollatos, O., Herbert, B. M., Matthias, E., & Schandry, R. (2007b). Heart rate response after emotional picture presentation is modulated by interoceptive awareness. *International Journal of Psychophysiology*, *63*(1), 117–124. https://doi.org/10.1016/J.IJPSYCHO.2006.09.003

Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion*, *18*(7), 1009–1023. https://doi.org/10.1037/emo0000413

Ring, C., & Brener, J. (2018). Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. *Psychophysiology*, *February*, 1–10. https://doi.org/10.1111/psyp.13084

Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: A cautionary tale about interoceptive awareness. *Biological Psychology*, *104*, 193–198. https://doi.org/10.1016/J.BIOPSYCHO.2014.12.010

Rohrmann, S., & Hopp, H. (2008). Cardiovascular indicators of disgust. *International Journal of Psychophysiology*, *68*(3), 201–208. https://doi.org/10.1016/j.ijpsycho.2008.01.011

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (Contempt, Anger... *Journal of Personality and Social Psychology*, *76*(4), 574–586. https://doi.org/10.1037/0022-3514.76.4.574

Salvato, G., De Maio, G., & Bottini, G. (2019). Interoceptive sensibility tunes risk-taking behaviour when body-related stimuli come into play. *Scientific Reports*. https://doi.org/10.1038/s41598-019-39061-0

Schandry, R. (1981). Heart Beat Perception and Emotional Experience. *Psychophysiology*, *18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Scheepers, D., de Wit, F., Ellemers, N., & Sassenberg, K. (2012). Social power makes the

heart work more efficiently: Evidence from cardiovascular markers of challenge and threat. *Journal of Experimental Social Psychology*, *48*(1), 371–374. https://doi.org/10.1016/j.jesp.2011.06.014

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality & Social Psychology Bulletin*, *34*(8), 1096–1109. https://doi.org/10.1177/0146167208317771

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2015). Landy and Goodwin (2015) confirmed most of our findings then drew the wrong conclusions. *Perspectives on Psychological Science*, *10*(4), 537–538. https://doi.org/10.1177/1745691615589078

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0007

Shenhav, A., & Mendes, W. B. (2014). Aiming for the stomach and hitting the heart: Dissociable triggers and sources for disgust reactions. *Emotion*, *14*(2), 301–309. https://doi.org/10.1037/a0034644

Sood, S., & Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organizational Behavior and Human Decision Processes*, *98*(2), 144–154. https://doi.org/10.1016/j.obhdp.2005.05.005

Spielberger, C. D., & Gorsuch, R. L. (1983). *The state trait anxiety inventory for adults: sampler set manual, test, scoring key*. Redwood City Calif: Mind Garden.

Starcke, K., Ludwig, A. C., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral judgment. *Judgment and Decision Making*, *7*(1), 61–68. https://doi.org/10.17185/duepublico/45052

Suzuki, A., Hirota, A., Takasawa, N., & Shigemasu, K. (2003). Application of the somatic marker hypothesis to individual differences in decision making. *Biological Psychology*, *65*(1), 81–88. https://doi.org/10.1016/S0301-0511(03)00093-0

Sze, J. A., Gyurak, A., Yuan, J. W., & Levenson, R. W. (2010). Coherence Between Emotional Experience and Physiology: Does Body Awareness Training Have an Impact? *Emotion*, *10*(6), 803–814. https://doi.org/10.1037/a0020146

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson Eduation UK.

Takahashi, T., Ikeda, K., Ishikawa, M., Kitamura, N., Tsukasaki, T., Nakama, D., & Kameda, T. (2005). Anxiety, reactivity, and social stress-induced cortisol elevation in humans. *Neuroendocrinology Letters*, *26*(4), 351–354.

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. Evans & K. Frankish (Ed.), *In Two Minds: Dual Processes and Beyond* (pp. 171–195). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.003.0008

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Tracy, J. L., Steckler, C. M., & Heltzel, G. (2019). The physiological basis of psychological disgust and moral judgments. *Journal of Personality and Social Psychology*, *116*(1), 15–32. https://doi.org/10.1037/pspa0000141

Turner, M. J., Jones, M. V., Sheffield, D., Slater, M. J., Barker, J. B., & Bell, J. J. (2013). Who thrives under pressure? Predicting the performance of elite academy cricketers using the cardiovascular indicators of challenge and threat states. *Journal of Sport and Exercise Psychology*, *35*(4), 387–397. https://doi.org/10.1123/jsep.35.4.387

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.4324/9781912282562

Van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P. C., Schulz, A., & Herbert, B. M. (2016). The water load test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PLoS ONE*, *11*(9), 1–14. https://doi.org/10.1371/journal.pone.0163574

Vega, S., Mata, A., Ferreira, M. B., & Vaz, A. R. (2020). Metacognition in moral decisions: judgment extremity and feeling of rightness in moral intuitions. *Thinking and Reasoning*, *0*(0), 1–18. https://doi.org/10.1080/13546783.2020.1741448

Vianna, E. P. M., & Tranel, D. (2006). Gastric myoelectrical activity as an index of emotional arousal. *International Journal of Psychophysiology*, *61*(1), 70–76. https://doi.org/10.1016/j.ijpsycho.2005.10.019

Vicario, C. M., Kuran, K. A., Rogers, R., & Rafal, R. D. (2018). The effect of hunger and satiety in the judgment of ethical violations. *Brain and Cognition*, *125*(February), 32–36. https://doi.org/10.1016/j.bandc.2018.05.003

Weiss, S., Sack, M., Henningsen, P., & Pollatos, O. (2014). On the interaction of self-regulation, interoception and pain perception. *Psychopathology*, *47*(6), 377–382. https://doi.org/10.1159/000365107

Werner, N. S., Duschek, S., Mattern, M., & Schandry, R. (2009). The relationship between pain perception and interoception. *Journal of Psychophysiology*, *23*(1), 35–42. https://doi.org/10.1027/0269-8803.23.1.35

Werner, N. S., Jung, K., Duschek, S., & Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology*, *46*(6), 1123–1129. https://doi.org/10.1111/j.1469-8986.2009.00855.x

Whitehead, W.E., Drescher, V. M., & Heiman, P. (1977). Relation of heart rate control to heartbeat perception. *Biofeedback and Self-Regulation*, *2*(4), 317–392.

Whitehead, William E., & Drescher, V. M. (1980). Perception of Gastric Contractions and Self-Control of Gastric Motility. *Psychophysiology*, *17*(6), 552–558. https://doi.org/10.1111/j.1469-8986.1980.tb02296.x

Wiens, S., Mezzacappa, E. S., & Katkin, E. S. (2000). Heartbeat detection and the experience of emotions. *Cognition and Emotion*, *14*(3), 417–427. https://doi.org/10.1080/026999300378905

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to Ventromedial Prefrontal Cortex Impairs Judgment of Harmful Intent. *Neuron*, *65*(6), 845–851. https://doi.org/10.1016/j.neuron.2010.03.003

Zeng, J., Zuo, X.-L., Li, Y.-Q., Wei, W., & Lv, G.-P. (2007). Tegaserod for dyspepsia and reflux symptoms in patients with chronic constipation: an exploratory open-label study. *European Journal of Clinical Pharmacology*, *63*(6), 529–536. https://doi.org/10.1007/s00228-007-0287-3

Study 2 Postface

Paper:

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). *Do individual differences in interoception influence the relationship between physiological arousal and moral judgments of harm? [Unpublished thesis]*. University of Bath.

Overall, an ability to detect heartbeat sensations had several interesting moderation effects on the relationship between cardiovascular arousal and harm aversion and outcome maximisation tendencies. It is possible that we did not find the direct relationships between the strength of physiological arousal and moral judgments as in prior psychophysiological studies (e.g. Moretto et al., 2010) because we used a fundamentally different moral dilemma paradigm that measures the relative motivation to avoid harm or maximise outcomes, as opposed to inferring motivation from moral judgments that are simultaneously influenced by both of these drives (Conway & Gawronski, 2013). These stimuli have revealed nuances in the socio-emotional motivations behind people's judgments that traditional sacrificial dilemmas could not (Reynolds & Conway, 2018). Thus, the findings clarify some of the relationships between physiological arousal and discrete motivational tendencies of harm aversion and outcome maximisation when making egocentric moral judgments, and for allocentric acceptability judgments of 'non-utilitarian' and 'utilitarian' harmful acts.

The moderating effect of interoceptive accuracy in the relationship between cardiovascular arousal and moral judgments applied mostly to egocentric moral judgments (harm-aversion and outcome-maximisation), suggesting that an ability to perceive one's heartbeats may only prove influential in decisions where we consider ourselves as being the agents of harm, which is consistent with related neuropsychological evidence. A neuroimaging study found the amygdala was activated only when participants considered their own transgressions of social norms, compared to someone else's transgressions, indicating activation of this area is modulated by intentionality and agency (Berthoz et al., 2006). The authors suggest the amygdala is implicated in emotional processes concerning personal welfare, as the transgressions are weighed up according to the self-relevant consequences. Afferent cardiac signals can directly influence activation of the amygdala,

which is key in the processing of threatening stimuli, and is sensitive to changes in autonomic arousal (Garfinkel & Critchley, 2016). For example, patients with amygdala damage also show deficits in their ability to perceive arousal based negative emotions (Adolphs et al., 1999). The Somatic Marker Hypothesis (Damasio, 1996) suggests that bodily states encoding reward versus punishment values of potential outcomes can bias decision-making under uncertainty. These bodily states are activated by areas including the amygdala to communicate a holistic representation of information about the bodies internal state, current goals, and the external context. The insula is believed to play a role in calculating risk associated with specific actions (Damasio, 1996; Naqvi, Nasir & Bechara, 2010); is centrally implicated in interoceptive processing (Craig, 2015; Seth & Friston, 2016); and is activated when people pay attention to internal bodily states (Critchley et al., 2004; Terasawa et al., 2013). In the current study, greater cardiovascular threat reactivity was associated with reduced harm aversion tendencies for people who could more accurately perceive their heart rate. For these people, weighing up the prospect of accepting harmful action on incongruent trials was associated with an enhanced sympathetic cardiovascular response, potentially generated by consideration of the negative consequences of causing harm for their personal welfare i.e. social condemnation and reputational damage associated with carrying out harmful acts. However, better heartbeat detectors may have a superior ability to integrate bodily feedback regarding self-relevant consequences of harming one person with information about the wider context of the dilemma, associated with risk assessment processes in the insula. This may have manifested in relatively reduced harm-aversive response tendencies, if they were less likely to reject harm on the basis of an aversive bodily state.

We speculated that the inverse relationship between cardiovascular threat reactivity and harm-aversion may be facilitated by a greater tendency among higher IAc people to exhibit perceptual inference in the presence of interoceptive prediction error (i.e. update the relative precision of the prior) when faced with surprising changes in interoceptive signals (Farb et al., 2015; Ainley et al., 2016). Heartbeat detection ability has been associated with a superior ability to enhance the precision of incoming interoceptive prediction errors compared to prior states in memory, as better heartbeat perceivers are able to use attention to prioritise interoceptive information over other sensory information, which results in a reduction in prediction error and energy usage over time (Ainley et al., 2016). Both active (updating interoceptive state to assimilate with prior expectation) and perceptual forms of

inference can be adaptive forms of interoceptive inference. However, Farb et al (2015) suggest there may be benefits to perceptual forms of inference for reducing the likelihood of overt behavioural responses to achieve sensory regulation, which facilitates greater 'non-reactivity' to surprising interoceptive information. We suggested this overt behavioural response in the context of moral dilemmas would be to reject harmful actions, in the presence of an aversive physiological response originating from behavioural conditioning linking harmful actions with aversive outcomes (Miller et al., 2014). Importantly, Farb et al (2015) suggest that beyond interoceptive sensitivity, goals of regulation versus accuracy will determine whether active or perceptual inference is overriding, but forms of perceptual inference may support greater reflection of the evolving time-course of arousal rather than facilitating a rapid regulatory response. Further work using EEG or fMRI methods is needed to understand the potential neural mechanisms underlying the inverse relationship between greater cardiovascular threat reactivity and a reduction in harm-aversion for better heartbeat detectors.

Another key finding was that lower heart rate on incongruent trials compared to congruent trials, was associated with outcome-maximisation tendencies, but only for people with superior heartbeat detection ability. This has implications for dual-process models (Greene et al., 2001, 2004), as it suggests that there is an emotional component to outcome-based moral judgments that is influenced by individual differences in an ability to perceive cardiac sensations. Evidence for an emotional component in outcome-based moral judgments is consistent with Reynold and Conway's (2018) finding that an affective concern for others was associated with the outcome-maximisation parameter using the same moral dilemma task. Heart rate deceleration is associated with negative emotional responses to aversive stimuli (Lang et al., 1993) including 'personal' moral dilemmas (Carmona-Perera et al., 2013). However, heart rate deceleration has also been associated with the facilitation of sensory processing (Lacey & Lacey, 1978) as a means to reduce afferent feedback associated with heartbeats that can distract from perceptual and cognitive processes (Garfinkel & Critchley, 2016). Therefore, lower heart rate on incongruent trials may have represented a functionally adaptive response to support visual and cognitive processing of the moral dilemma stimuli. Subsequently, lower heart rate may have been positively associated with outcome-maximisation response tendencies, if greater sensory processing of the moral dilemmas encouraged people to fully consider the context of the dilemma, potentially 'overriding' a negative emotional response to the prospect of accepting harm and choosing a

response that benefitted the most people. Conversely, greater heart rate also predicted an increase in harm-aversion tendencies for better heartbeat detectors. A link between heart rate acceleration and harm-aversive responses suggests that moral judgments on incongruent trials were potentially preceded by shallower sensory processing of the moral dilemma stimuli, possibly resulting in more harm-rejection responses. Further work is needed to clarify the roles of emotional experience and sensory processing in the link between anticipatory heart rate deceleration and acceleration and moral judgments of harm.

The discrete influences of the physiological parameters on moral judgments suggests a 'selectivity' ability among better heartbeat perceivers, in weighing up different types of bodily feedback when forming moral judgments and could be an important consideration for future psychophysiological studies using this paradigm. Heart rate remains an ambiguous parameter of emotional arousal which warrants further investigation in this context. Overall, capturing a range of physiological measures has shown to be useful for understanding how a constellation of psychophysiological events may offer unique contributions to moral judgments for people who are better at perceiving cardiac sensations.

Study 3 Preface

Paper:

Brown, H., Fraser, D. S., Brown, S. G. R., & Proulx, M. J. (2020). *Moral dilemmas in a semi-autonomous vehicle: the effects of physiological arousal, interoception and typical versus non-typical driving actions in VR driving dilemmas [Unpublished thesis]*. University of Bath.

Pre-registration:

Brown, H., Fraser, D. S., & Proulx, M. J. (2020a). *Does modality of harmful action in a semi-autonomous vehicle influence responses to virtual-reality moral dilemmas?* Open Science Framework. https://osf.io/vhbdy.

Draft paper to be submitted to:

IEEE-VR

The primary aim of study 3 was to explore the role of interoception in moral dilemma situations where the participant is physically immersed in a sensory environment and their body is physically involved in carrying out harmful acts that could generate stronger aversions to action-based harm (Cushman et al., 2012; Francis et al., 2017). Ultimately, we wanted to understand whether interoception interacted with physiological arousal to influence moral behaviour in a different way to moral decision-making found in study 2. This was largely driven by recent VR work (e.g. McDonald et al., 2017; Patil et al., 2014), showing that people experience strong emotional and/or physiological response in virtual reality moral dilemmas which can predict the likelihood that they will carry out harm. VR research in this field has also revealed discrepancies between what people say they would do when responding to a text-based moral dilemma, and what they actually do when faced with similar scenarios in VR (Francis et al., 2017; Patil et al., 2014).

The context of semi-autonomous vehicles presents a modern-day moral dilemma in terms of the prospective, automated morality algorithms currently being developed for these vehicles. But, more urgently, there is a need to better understand how humans may behave in morally sensitive driving situations where they are required to 'rubberstamp' or take control of automated functions that could have implications for the wellbeing of others. In the near

future, drivers are likely to be required to provide oversight to vehicle AI in a way that will fundamentally change the driving experience, and our sense of responsibility and agency in the driving seat (Limerick et al., 2014). Driving dilemmas also provide a useful, and more ecologically valid context than traditional Trolley and Footbridge dilemmas to expose participants to multiple 'plausible' and novel moral dilemma scenarios. We used matched congruent and incongruent moral dilemmas, to explore a scaled VR version of Conway and Gawronski's (2013) procedure for exploring harm-aversion and outcome-maximisation tendencies. In the incongruent dilemmas, participants had to decide whether to harm one person (a pedestrian) in the path of the vehicle, in order to save more people in the vehicle who would be fatally harmed in an unavoidable collision (incongruent). In the congruent dilemma, the context was the same but harming the pedestrian only resulted in avoiding damage to the vehicle, therefore, a motivation to maximise outcomes would not be driving choices to harm the pedestrian. Again, this allowed us to explore differences in physiological arousal in moral dilemma conditions where there is a conflict between motivations to maximise outcomes (utilitarian) and avoid harm (deontological), and conditions where there is no motivation to maximise outcomes.

We were also interested in whether action-based harm-aversion could predict different physiological and behavioural responses for typical (an accelerator foot-pedal) versus non-typical (a button-press) harmful actions. These actions do not involve direct contact with another person but instead require carrying out an action in a vehicle the participant is sitting in to cause harm, whereas previous studies exploring action-based harm aversion have used simulated actions such as hitting, cutting and hammering another person (Cushman et al., 2012; Parton & McGinley, 2019). We wanted to explore whether the directness of harm (Greene et al., 2001) could extend to actions that use the agents' body but are not physically touching the victim, with the foot-pedal condition potentially resembling a more aversive action, due to learned associations with pressing an accelerator pedal when driving. The dilemmas were presented in the same order, as order effects have been found for judgments in text-based moral dilemmas if a more arousing dilemma is presented after a less arousing dilemma (Schwitzgebel & Cushman, 2012).

We expected that individual differences in interoception would facilitate a different relationship between moral behaviour and arousal, compared to study 2, due to fundamental differences in their role in the task. The egocentric perspective when carrying out harmful

actions (Francis et al., 2017) and the contextual saliency of the potential consequences of harm are enhanced in VR dilemmas (Patil et al., 2014). As the self-relevant consequences of harmful action are clearer, this could lead to greater activation of brain areas implicated in emotional processes regarding personal welfare (Berthoz et al., 2006). Here the participants were 'drivers' in a semi-autonomous vehicle where the choice to harm a pedestrian outside of the vehicle had consequences for themselves and other people in the vehicle. Therefore, their 'role' in carrying out harmful action was much more socially visible, but they also had a more immediate 'stake' in the outcome of their decision, as they could see the impending collision with another vehicle that required a rapid response. All these things considered, we expected interoceptive processes would be influential in situations where the [hypothetical] physical integrity of participants bodies were on the line. In addition, due to the fundamental associations proposed to exist among experiences of conscious presence, interoception and agency (Seth et al., 2012) we wanted to investigate the relationship between interoception, arousal, presence and virtual-reality sickness in VR moral dilemmas, as a novel investigation in this context.

Statement of Authorship

This declaration concerns the article entitled:			
Moral dilemmas in a semi-autonomous vehicle: the effects of physiological arousal, interoception and typical versus non-typical driving actions in VR driving dilemmas			
Publication status (tick one) **Draft manuscript** [x] **Submitted** [ ] **In review** [ ] **Accepted** [ ] **Published** [ ]			
Publication details (reference)	Draft paper to be submitted to: *IEEE Virtual Reality Conference Proceedings* Authors: Helen Brown, Danaë Stanton Fraser, Stephen G. R. Brown, and Michael J. Proulx.		
Copyright status (tick the appropriate statement) I hold the copyright for this material [X] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed the formulation of ideas, design of methodology, experimental work and presentation of data in journal format. The candidate (H.B.) conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper. S.G.R.B. contributed to the design and building of the virtual reality environments.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	This information has been redacted for privacy reasons.	**Date**	05/11/2020

Moral dilemmas in a semi-autonomous vehicle: the effects of physiological arousal, interoception and typical versus non-typical driving actions in VR driving dilemmas

Helen Brown [1,2], Danaë Stanton Fraser [2], Stephen G. R. Brown [3], and Michael J. Proulx [1]

[1] Crossmodal Cognition lab, Department of Psychology, University of Bath, Bath, UK

[2] CREATE Lab, Department of Psychology, University of Bath, Bath, UK

[3] Immersive Media Observatory, College of Engineering, Swansea University, Swansea, UK

Author Note

Research funded by University of Bath. This studentship runs alongside UK EPSRC grant Virtual Realities: Immersive Documentary Encounters (ref: EP/P025595/1).

Correspondence concerning this article should be addressed to Helen Brown, hrhb20@bath.ac.uk.

Abstract

In the near future, humans will be required to oversee morally sensitive decisions in semi-autonomous vehicles that have implications for public safety. Greater psychological distance from immoral acts can minimise the emotional responses towards them (Eyal & Liberman, 2012). In particular, more directly harmful acts are seen as less acceptable and more emotionally aversive than more indirect acts that achieve a utilitarian outcome (Cushman et al., 2006; Greene et al., 2001). Thus, we suggest how drivers interact with non-typical automation in semi-autonomous vehicles to execute decisions in emergency situations could reduce aversive responses to carrying out potentially harmful acts. This initial study (n=25) explored whether carrying out harm with a more typical driving action (pressing a foot-pedal), compared to a less typical action (pushing a button) influenced physiological arousal, willingness to carry out harm and response time to carry out harm in VR moral dilemma driving scenarios. Interoception (i.e. perception of visceral sensations) was also measured, due to fundamental associations with emotional experience and physiological reactivity (Craig, 2015; Füstös et al., 2013; Pollatos et al., 2007b). Greater electrodermal activity was associated with longer response times to carry out harmful action. The relationship between cardiovascular reactivity (pre-ejection period) and response time differed between the foot-pedal and button-press conditions and appeared to be associated with interoceptive accuracy. An ability to perceive one's heartbeats (interoception) may influence the regulation of aversive physiological states in moral dilemma scenarios where harmful action must be performed to achieve a utilitarian outcome. These preliminary findings highlight the need to consider interoceptive processes in higher-powered moral dilemma studies.

Keywords: virtual reality, VR, moral dilemmas, interoception, arousal, physiology, autonomous vehicles, harm, moral behaviour, moral decision-making, heartbeat detection.

Introduction

The role of the body in moral decision-making

The body plays an important role in how we feel an aversion to condoning harmful acts (e.g. Greene et al., 2001) and in how the sensorimotor qualities of harmful acts become laden with aversive qualities themselves, independent of their outcome (Cushman et al., 2012; Miller et al., 2014; Parton & McGinley, 2019). Greene and colleagues' (2001) influential dual process model emphasised how strong emotional reactions to the prospect of physically harming others can lead people to reject harm in hypothetical moral dilemmas, even when harm serves a 'greater good'. An aversion to action-based harm has been associated with individual differences in indexes of sympathetic cardiovascular arousal (Cushman et al., 2012; Parton & McGinley, 2019). Importantly, individual differences in the perception of visceral sensations, like heartbeats, is likely to be important in the link between physiological arousal and moral decision-making (e.g. Brown, Proulx, & Fraser, 2020; Dunn et al., 2010; Gu et al., 2013; Schnall et al., 2008). Interoception refers to the perception of the current physiological state and processes of our body and is fundamentally connected with our autonomic nervous system (Craig, 2015). Individual differences in interoception may characterise the relationship between body and mind (Feldman Barrett et al., 2004; Farb et al., 2015; Pollatos et al., 2007a, 2015) and shed light on the relationship between physiological arousal and harm-based moral decision-making.

The Somatic Marker Hypothesis (SMH; Bechara et al., 2005; Damasio, 1996) was the first neuropsychological theory to illuminate how the processing of physiological signals in the ventromedial pre-frontal cortex (VMPfC) could bias moral decision-making (Damasio et al., 1990; Koenigs et al., 2007; Young et al., 2010). Damage to the VMPfC has been associated with greater harm-acceptance in moral dilemma paradigms (Moretto et al., 2010) and deficits in social emotions (Anderson et al., 2013). The centrality of emotional, physiological, and bodily factors in harm-based moral decision-making highlights the limitations of 'cold' lab-based moral judgment research for understanding more realistic moral decision-making. Researchers are increasingly capitalizing on virtual reality (VR) technology to investigate moral behaviour in more immersive moral dilemma scenarios (Francis et al., 2016; Pan & Slater, 2011; Patil & Silani, 2014) including moral dilemma driving scenarios (Uijong et al., 2019), that would be ethically impossible in the real-world. Perhaps unsurprisingly, people are more likely to carry out harmful actions in VR sacrificial

dilemmas (where one person must be sacrificed to save several others), despite condemning the same action in text-based dilemmas (Francis et al., 2016; Patil et al., 2014). Lab-based moral judgment research often gives way to lengthy reasoning processes not present in real-life situations (Haidt, 2001). Our emotional reactions tend to be greater for events closer in time and space and when events happen to us rather than others (Eyal & Liberman, 2012; Trope & Liberman, 2010).

In sacrificial VR dilemmas, the enhanced contextual saliency of the negative outcomes of *not* causing harm can surpass the negative value associated with carrying out a harmful act (Cushman, 2013; Patil et al., 2014). Furthermore, the embodied and egocentric perspective of participants (Francis et al., 2017; Sood & Forehand, 2005; Tassy et al., 2013), physical agency to carry out 'harmful' actions (Francis et al., 2017), and self-preservation motivations in situations of 'self-sacrifice' (Faulhaber et al., 2018; Uijong et al., 2019), may be key factors influencing emotional experience and moral behaviour. In particular, VR offers a useful methodology for investigating moral behaviour in semi-autonomous (SA) vehicles. This new technology is fundamentally changing the driving experience, and in the near future these vehicles are likely to require human supervision in dynamic, morally sensitive situations such as an unavoidable collision. How drivers interact with SA vehicles in instances representing real-life moral dilemmas, could have serious implications for their behaviour and the wellbeing of pedestrians and other drivers.

What stops us from carrying out harmful acts?

Historically, moral dilemma research has focused heavily on sacrificial dilemmas like the Trolley and Footbridge problems. The Trolley dilemma (Thomson, 1985) asks whether it is morally acceptable to switch a driverless trolley travelling towards five people (who would be killed) to an alternate track with one person on it. Whereas the Footbridge dilemma (Foot, 2003) asks whether it is acceptable to push a large man off a footbridge to stop the trolley from killing the five people but sacrificing the man in the process. Utilitarianism would see harm as being morally acceptable (Mill, 1998), as it maximises good for the most people, whereas deontological ethics would disagree, as this would undermine the rights of the individual harmed (Kant, 2018). The net-outcome of causing harm in the Trolley and Footbridge problems is the same, yet people are more accepting of pulling the switch than pushing the man (Greene, 2009; Greene et al., 2001; McDonald et al., 2017). Greene et al (2001) found that the Footbridge dilemma evokes a negative emotional response not elicited

by the Trolley dilemma, and utilitarian judgments showed patterns of brain activation in regions associated with cognitive control (Greene et al., 2004). Greene's dual-process model (2001) proposed that deontological judgments, were the result of a strong emotional response to the prospect of harming someone, or a reduced ability to engage in cognitive deliberation, and has gained considerable empirical support (e.g. Baron et al., 2015; Byrd & Conway, 2019; Cushman et al., 2012; Patil et al., 2020; Starcke et al., 2012).

People typically rate harmful actions that involve physical contact as less morally acceptable than harmful actions without physical contact (Cushman et al., 2006). However, the effect of 'personal force' i.e. using the body to inflict harm, has shown to only influence moral judgments in the absence of an 'intention' of carrying out harm (Greene et al., 2009). Outcome-aversion and action-aversion are proposed to be two distinct motivations deterring us from carrying out harmful actions. These aversions are proposed to represent other-focused (outcome) versus self-focused (action) emotional drivers that influence our choices to accept or reject harm (Miller et al., 2014; Reynolds & Conway, 2018). Outcome-aversion is associated with a negative reaction to witnessing real-harm even if we are not the cause. In contrast, action-aversion represents our negative emotional response to carrying out prototypically harmful actions independent from our emotional response to the consequences of harm. Following consistent pairings of harmful actions (e.g. punching someone) with victim distress (e.g. facial expressions indicating pain, fear) the motoric and perceptual features of harmful actions become laden with aversive qualities (Miller et al., 2014) that can generate a negative emotional response (Blair, 1995; Cushman et al., 2012) which can predict moral judgments of personal harm (Cushman et al., 2012). Moreover, individual differences in self-reported action-aversion and outcome-aversion have shown to predict the rejection of harm in text-based moral dilemmas (Miller et al., 2014; Reynolds & Conway, 2018).

Harm rejection judgments are typically associated with aversive anticipatory physiological responses to harmful action, including skin-conductance (McDonald et al., 2017; Moretto et al., 2010) and heart rate deceleration (Carmona-Perera et al., 2013). In particular, measures of cardiovascular sympathetic reactivity including systemic vascular resistance (Cushman et al., 2012) and pre-ejection period (Parton & McGinley, 2019) are associated with an action-based aversion to harm. Systemic vascular resistance (SVR) is measure of resistance in the circulatory systems that corresponds with the release of epinephrine and the tightening of blood vessels and is associated with activation of the

sympathetic nervous system and psychological threat states (Blascovich & Mendes, 2000). Whereas challenge states are associated with adaptive increases in cardiac output and heart rate (Blascovich & Mendes, 2000; Scheepers et al., 2012). Pre-ejection period is a sympathetically innervated measure of cardiac contractility (Newlin & Levenson, 1979) and is proposed to be a more appropriate measure of sympathetic arousal than SVR, in situations that require active [as opposed to passive] coping- characterised as the ability to influence the outcome of an event through mental or physical effort (Parton & McGinley, 2019; Sherwood et al., 1990). Decreases in PEP when carrying out prototypically harmful actions are associated with increased sympathetic activation, attributed to the effort required to make the action and the aversive nature of the harmful act (Parton & McGinley, 2019).

Interoception and psychophysiological arousal

In a given context, our emotional responses and behavioural motivations towards environmental stimuli will be generated by interoceptive processes, influenced by the body's current state (Feldman Barrett & Simmons, 2015). Afferent heartbeat signals can support the processing of threatening stimuli (Garfinkel & Critchley, 2016) and even the belief of heightened cardiovascular arousal can influence moral decision-making (Gu et al., 2013). However, the role of interoceptive capacities in the relationship between physiological arousal and moral behaviour has not previously been explored. Interoception is multi-dimensional and can be assessed across visceral systems in the body. Interoceptive dimensions have shown to be distinct (Garfinkel et al., 2015) and may be associated with moral judgments in different ways (Brown, Proulx, & Fraser, 2020; Brown, Proulx, & Fraser, 2020b).

Interoceptive accuracy (IAc), refers to an ability to accurately perceive or detect internal sensations such as heartbeats (Schandry, 1981), whereas interoceptive sensibility (IS) refers to a self-reported tendency to focus on or pay attention to bodily sensations, such as heartbeats, or changes in body temperature (Garfinkel et al., 2015; Miller et al., 1981). Interoceptive meta-cognitive awareness (IAc-meta) captures someone's awareness of their interoceptive accuracy which can be measured as how confident people are that their perceptions of internal sensations, such as heartbeats, accurately reflect real visceral events verified by physiological recording devices (e.g. Garfinkel et al., 2016). Differences in sensitivity to gastric processes like hunger have been linked to moral judgments in prior research, potentially due to a link with disgust emotions (e.g. Brown, Proulx, & Fraser,

2020b; Schnall et al., 2008; Vicario et al., 2018), but individual differences in gastric interoception has not previously been explored in studies of moral behaviour.

Interoceptive accuracy in particular has been associated with more intense emotional experiences (Wiens, 2005) and psychophysiological reactivity to physical stress (Eichler & Katkin, 1994; Pollatos et al., 2007a) and emotional stimuli (Pollatos et al., 2007b). Pre-ejection period (PEP) in particular has been associated with heartbeat detection and represents the strength of heart contractions, with shorter PEP indicating a stronger force of myocardial contraction (Eichler & Katkin, 1994). Katkin (1985) proposed that when the heart contracts with greater power, this may give a more pronounced visceral "thump" sensation which contributes to better heartbeat detection. In support, Eichler et al (1987) found PEP was greater for better heartbeat detectors when exposed to aversive stimuli. However, trait anxiety has also been associated with heightened physiological reactivity to stressors (Pollatos et al., 2007a; Takahashi et al., 2005) and ability to perceive heartbeats (Critchley et al., 2004; Pollatos et al., 2007a). People who typically focus on visceral sensations may be more likely to perceive changes in physiological arousal and misinterpret these changes as meaningful (Clark et al., 1997; Domschke et al., 2010), which could potentially lead to increases in anxiety (Paulus & Stein, 2010). This also relates to the dimension of interoceptive sensibility; as self-reported hypervigilance to somatic sensations has been implicated in a range of anxiety disorders (e.g. Anderson & Hope, 2009; De Berardis et al., 2007).

A recent study found IAc moderated the relationship between cardiovascular threat-reactivity and harm-aversion tendencies in a moral dilemma task (Brown, Proulx, & Fraser, 2020b). Therefore, although IAc may predict greater physiological reactivity towards stressors, higher IAc people may be less 'behaviourally reactive' towards aversive physiological sensations when making moral judgments (Brown, Proulx, & Fraser, 2020b). Predictive coding models of interoception suggest that when there is a mismatch between expected and actual (or desired) sensory information i.e. prediction error, comparisons are made between current sensations and simulated sensations in the past and future (Apps & Tsakiris, 2014; Farb et al., 2015; Friston et al., 2010; Seth & Friston, 2016). Larger magnitudes of prediction error are more likely to reach conscious awareness and drive a regulatory response. These comparisons can lead to active and perceptual forms of inference that motivate behaviour to resolve the conflict.

Perceptual inference prioritises accuracy and precision of sensory information i.e. the expected state is altered to be in-line with the current sensory state. Whereas, less accurate interoceptive representations may be associated with active forms of inference that prioritise sensory regulation, where the sensory state is changed to fit with the expected state (Apps & Tsakiris, 2014; Farb et al., 2015). Farb and colleagues (2015) propose perceptual inference is likely to be associated with more accurate interoceptive representations and may reduce the power of strong, surprising interoceptive signals to influence overt behaviour. Relatedly, Ainley et al (2016) suggest that people who are able to direct attention to their heartbeats on a heartbeat detection task, must also have the ability to prioritise interoceptive signals over other sensory modalities and therefore higher IAc people would generate more precise interoceptive predictions. Whereas, for people who are poorer at perceiving heartbeats, the sensory modality of interoception would be less routinely salient (Ainley et al., 2016).

It is possible that a superior sensitivity towards interoceptive states may support more adaptive regulation of emotional responses as changes in physiological sensations can be perceived more accurately (e.g. Füstös et al., 2013). Interoceptive accuracy has been associated with more adaptive self-regulatory abilities in a range of domains, including self-control and down-regulation of cravings (Kruschwitz et al., 2019), pain-tolerance (Weiss et al., 2014), regulation of exercise (Herbert et al., 2007) and aversive states associated with social exclusion (Pollatos et al., 2015). Moreover, IAc has been positively associated with antecedent and response-focused emotional regulation strategies and downregulation of negative affective states (Füstös et al., 2013; Kever et al., 2015). We speculate that interoceptive accuracy may support perceptual forms of inference and adaptive emotional regulation strategies which could moderate the relationship between aversive physiological reactions toward harm, and harmful 'utilitarian' actions in a VR moral dilemma task.

Interoception is also closely tied to conscious experiences of presence in and outside VR, as predictions about our sensory experiences are mismatched with our current sensory experiences in virtual environments (Diemer et al., 2015; Seth et al., 2012). Presence refers to subjective experiences of 'being there' in a virtual environment (Slater & Wilbur, 1997). Interoceptive sensibility has recently been positively linked with experiences of presence in VR and negatively associated with motion-sickness, suggesting that attending to bodily sensations may improve feelings of spatial presence and reduce sensory-mismatch issues associated with cybersickness (Farb et al., 2015; Heeter et al., 2020). Diemer and colleagues

(2015) propose that cognitive appraisals of presence are partially due to the level of emotional arousal experienced and attributed to VR environments. It is currently unclear how individual differences across other interoceptive dimensions, alongside changes in physiological arousal may contribute to experiences of presence in VR moral dilemma.

Action aversion: a semi-autonomous vehicle scenario

As the sensorimotor properties of harmful acts have been shown to be important for evoking an emotionally aversive response (Cushman et al., 2012; Miller et al., 2014), we hypothesised that typical actions required to carry out harm in a semi-autonomous (SA) vehicle may be more effective at triggering an aversive response than others. As drivers, we associate the typical action of pressing an accelerator pedal with our foot, with a vehicle moving forward at increasing speed. In addition, cultural references of dramatic car accidents or collisions in films, tv, and books, and even personal experiences, pave the way for learned associations between pushing an accelerator pedal downwards and hitting another object such as a wall, another car, or even a person. As a comparison, we have not built up the same associations between pushing an arbitrary button on the dashboard, with accelerating forwards or any other consistent outcome as this would represent (for the majority of cars) a novel driving action.

In particular, buttons, such as doorbells or light switches, were originally designed to be every-day objects, effortless to operate. They simplified labour or domestic tasks by making the interaction between human and machine more “automatic”- hiding the complex processes of the technology (Plotnick, 2018). Whereas, an accelerator foot-pedal in a vehicle alters speed to varying degrees of pressure and is a more context-specific interaction. The design of cars, their functions and how we drive them is likely to change with the development of autonomous vehicles, potentially creating more distance between the driving acts we perform and the consequences of these actions. Psychological distance from immoral acts can influence moral judgments of these acts (Eyal & Liberman, 2012). In the Trolley dilemma, pulling a lever may feel like a relatively innocuous act, that creates both physical and psychological distance from the harmful consequences, whereas in the Footbridge dilemma, there is no escaping the direct association between pushing the man, and the harm caused to him (e.g. Francis et al., 2017). Novel or non-typical actions [that may cause harm] are likely to create more distance from the harmful nature of the act and could have implications for driving behaviour. We explored whether the nature of a harmful action in a

vehicle influences physiological responses to harmful actions and moral behaviour in a collision situation.

Present study

We investigated whether drivers of a semi-automated vehicle in VR, chose to 'override' the car's automatic braking function to harm a pedestrian in the path of the vehicle in order to avoid a collision that would kill all passengers in the vehicle (Figure 1). We tested whether physiological responses of participants predicted harm-avoidance behaviour or response time, and whether this differed depending on the harmful action-type. Harmful action type was either typical (foot-pedal) or non-typical (pushing a button). We hypothesised the typical 'foot-pedal' condition would be more physiologically aversive. A primary aim of this study was to explore individual differences in interoception in the relationship between physiological arousal and moral behaviour. The study protocol was pre-registered on the Open Science Framework (Brown, Fraser, & Proulx, 2020a; https://osf.io/vhbdy). Due to the onset of coronavirus, the planned study was relegated to a preliminary study and all hypotheses are presented as exploratory.

Exploratory hypotheses

H1. There will be an effect of dilemma-type (congruent versus incongruent), timeframe (pre-warning, pre-decision, post-decision) and action-type (button-press versus foot-pedal) on physiological arousal.

H2. There will be an interaction between physiological arousal and action-type for predicting moral behaviour in the VE.

H3. Interoceptive accuracy (IAc) will be associated with greater cardiovascular reactivity on incongruent trials.

H4. Interoception will be associated with experiences of presence in the VE.

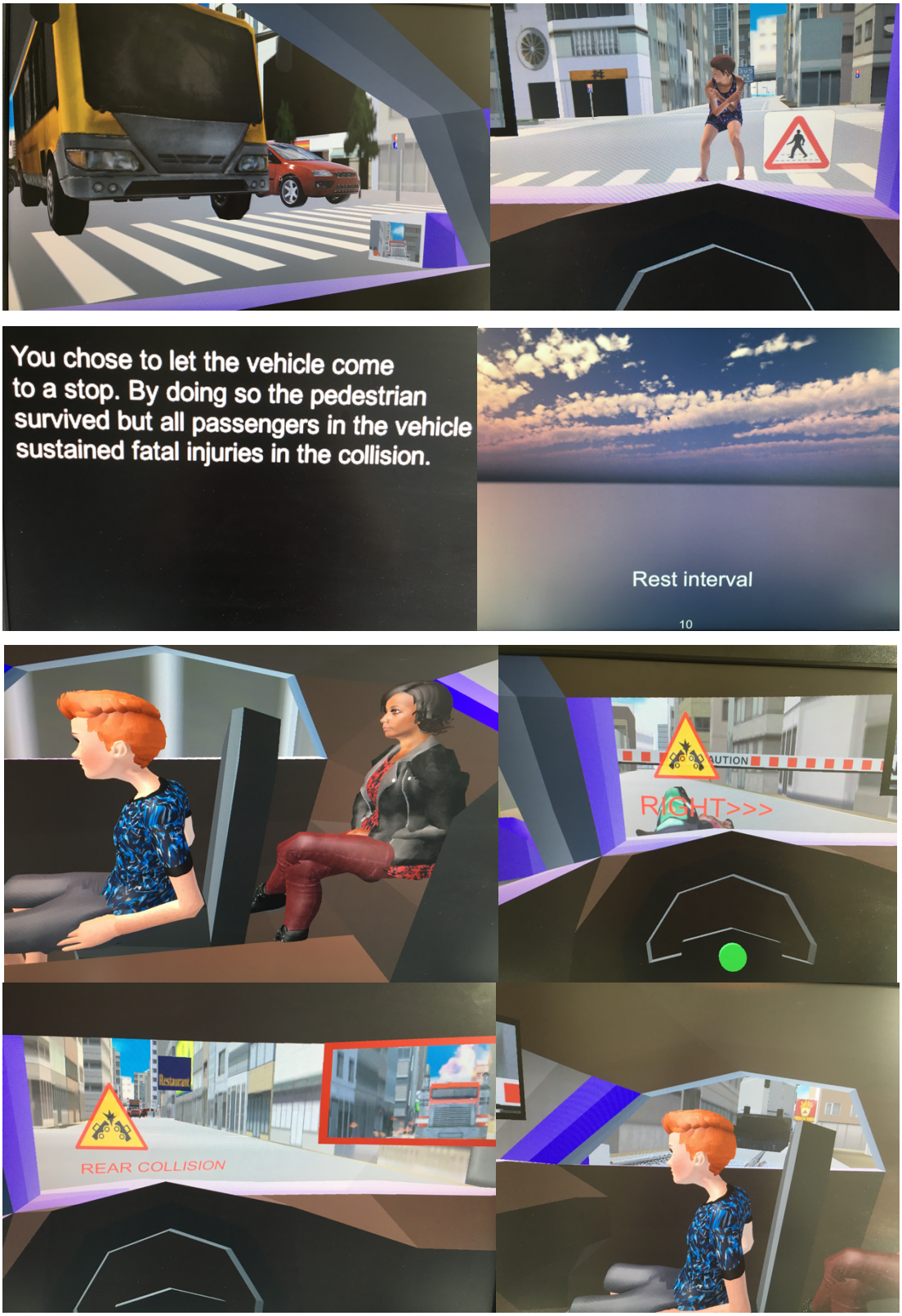
H5. Presence will be associated with greater difference in physiological arousal on incongruent trials.

H6.1. There will be an interaction between PEP reactivity and action-type for predicting response time for harmful action.

H6.2. Interoceptive accuracy will be associated with action-type and response time.

Figure 1.

Screenshots of the VR moral dilemma scenes. 1. Bus collision dilemma 2 (top left), 2. Pedestrian dilemma 1 (top right), 3. Text depicting outcome (below 1.), 4. Rest screen (below 2.), 5. Passengers (below 3), 6. Collision warning dilemma 3 (below 4), 7. Collision warning dilemma 1 (below 5), 8. Oncoming train collision dilemma 3 (below 6)



Method

Design

This was a 2 (dilemma type: congruent versus incongruent) x 2 (action type: button-press versus foot-pedal) mixed design (n=25). Action type was the between-subjects factor. Interoception, and physiological arousal during the moral dilemma task were the primary independent variables. Physiological arousal, harmful action on congruent and incongruent trials and response time for harmful action were the primary dependent variables.

Participants

Ethics approval was received from University of Bath ethics committee. Participants were recruited from University of Bath Student Participant Pool, the Psychology Department community panel database and departmental email lists. Subjects received course-credit or £10-15 in exchange for their time. Using G*power ($\alpha = 0.95$, $\beta = 0.8$, d=.05), sample size recommendation was 51 participants per condition (n=102), although 40 per condition was considered a more feasible aim. We achieved a total sample of 25 participants (button-press n=19, foot-pedal n=6), due to the onset of the coronavirus pandemic in March 2020. The sample was 64% female, with age range 18-28 (Median=20, SD=2.78). Healthy righthanded people aged 18-50 with driving experience were recruited. Participants were ineligible to take part if they were pregnant; had heart/gastrointestinal surgery or condition; were experiencing mental health issues; history of renal problems; had experienced a traumatic road incident in the past; had physical conditions or medication that affected diet or weight or ability to safely take part in the study. A virtual reality health and safety checklist was also completed by participants to assess any health risks associated with the technology.

Questionnaires

Anxiety

The State and Trait Anxiety Scale (Spielberger & Gorsuch, 1983) asks people to indicate their agreement (Not at all/Somewhat/Moderately so/Very much so) with twenty positively and negatively coded items e.g. *'I feel calm', 'I feel tense', 'I feel at ease'*. There are two identical scales that ask people to report how they feel in the current moment (State

anxiety) and how they feel in general (Trait anxiety). Coded responses are added together to calculate measures of State and Trait anxiety.

Interoceptive sensibility

Interoceptive sensibility was measured using the Private Body Consciousness (PBC) subscale of The Body Consciousness Questionnaire (BCQ: Miller et al., 1981). The complete BCQ was used to maintain scale-validity. The PBC subscale includes 5 statements: *'I'm very aware of changes in my body temperature', 'I am quick to sense the hunger contractions in my stomach', 'I know immediately when my mouth or throat gets dry', 'I am sensitive to internal bodily tensions',* and '*I can often feel my heart beating'*. Participants are asked to rate how characteristic each item is of themselves (extremely uncharacteristic/ uncharacteristic/ neutral/characteristic/extremely characteristic). Scores were numerically coded 1-5 (1=extremely uncharacteristic) and a mean score calculated.

Presence

Subjective experiences of presence in the VR environment (VE) were measured using an adapted SUS (Slater-Usoh-Steed) VR Presence scale (Slater et al., 1995, 2000). The scale included six items that capture 1) the sense of physically 'being in' the VE, 2) the extent the VE becomes the prevailing reality and 3) how much the VE resembles a memory of a "place". We adapted the scale questions for our VR environment e.g. "*The car and city seem to me to be more like...*" (0= 'Images that I saw', 7= 'Somewhere that I visited'), and "*I had a sense of 'being there' in the car and city*" (0= 'At no time', 7= 'Almost all the time'). The scoring method for SUS proposes a response variable calculation of the number of answers that were a rating of 6 or 7. The small sample size meant the proportion of people reporting scores of 6 and 7 was relatively low, therefore, we also calculated a mean score from the 6 items to represent a continuous measure of presence.

Virtual reality sickness

As interoceptive awareness has shown to be negatively associated with motion sickness (Heeter et al., 2020), we included a measure of virtual reality sickness to explore associations with the interoception variables. The virtual reality sickness questionnaire (VRSQ) (Kim et al., 2018) was used. The scale assesses oculomotor (e.g. fatigue, eyestrain) and disorientation (e.g. blurred vision, headache) experiences. People are asked to rate how

much they experienced each symptom during the VE (not at all, slightly, moderately, very). Oculomotor and disorientation symptoms are computed separately (Oculomotor ([1]/12)*100, Disorientation ([2]/15)*100), and a total score calculated (Oculomotor score + Disorientation score)/2.

Interoception tasks

Interoceptive accuracy and meta-cognitive awareness

Cardiac interoceptive accuracy was measured using procedures from Schandry's (1981) heartbeat mental tracking task. A pulse oximeter (BLYL CMS 50D+) was attached to the index finger on the right hand. Participants were instructed to silently count their heartbeats for four trial periods; 30, 50, 40, and 20 seconds (in the same order), while seated. The beginning of each trial was indicated by the experimenter and an alarm signalled the end of each trial. Participants reported how many heartbeats they perceived for each trial and reported their confidence in the accuracy or their estimation i.e. *How sure are you that the amount of heartbeats you counted for that interval were correct?* (0=not sure at all, 8= absolutely sure). Within-subjects' correlations (Pearson's *R*) of tracking accuracy and confidence ratings provided the interoceptive meta-cognitive awareness measure. Interoceptive accuracy was calculated using the formula below (Garfinkel et al., 2015; Hart et al., 2013):

$$1- \frac{|\, nbeats_{real} - nbeats_{reported} \,|}{(nbeats_{real} + nbeats_{reported})/2} \;;$$

Gastric interoception

Gastric interoception was measured using a two-step Water-Load Test (Van Dyck et al., 2016) to capture interoceptive sensitivity to gastric processes. Participants were sat upright and drank non-carbonated tap water at room temperature from a 400ml opaque flask with a straw. They were asked to drink freely until 1) they noticed sensations of stomach bloating or distention (Zeng et al., 2007), and 2) when their stomachs were completely full of water, at which time they notified the experimenter and the water consumed was measured. Participants were expressly reminded they could stop at any time and to notify the experimenter if they felt unwell. Maximum drinking time was ~5 minutes. The maximum

amount of water available was 1300ml, but participants were provided with water in 200ml amounts to blind them to the amount they were drinking. We added 2.5ml of electrolyte solution to the water to mitigate against a loss of sodium in the blood. Gastric interoception was calculated as the percentage of water drunk from the first signs of stomach distension at step 1, to achieve maximum fullness at step 2- providing an index of gastric sensitivity not dictated by stomach capacity (Van Dyck et al., 2016). This measure indicates the onset of visceral awareness from the first signs of stomach distension to complete fullness. Lower scores indicate an increased sensitivity to gastric processes earlier in the stomach-filling process. As a manipulation check, a short hunger and thirst measure were completed before and after the task: *How hungry do you feel at this moment?* (1=not at all, 9=extremely hungry), *How thirsty do you feel at this moment?* (1=not at all, 9=extremely hungry).

Moral dilemma task

Moral dilemma stimuli

The moral dilemma stimuli involved six moral dilemmas (3 x incongruent and 3 x congruent that were matched) for each participant. Incongruent dilemmas involved a situation where harm (killing a pedestrian) was necessary to maximise outcomes (save all passengers in the vehicle), whereas congruent dilemmas involved a situation where harm resulted in trivial outcomes (e.g. damage to vehicle), as described by Conway and Gawronski's (2013) text-based moral dilemma procedure. This was to assess whether we could calculate the probability people would avoid harm, while controlling for their tendency to maximise outcomes. Each dilemma involved an unfolding collision event with a vehicle from behind (Dilemma 1), a vehicle approaching from the left (Dilemma 2) and a vehicle approaching from the right (Dilemma 3) in three different locations around a VR city. Each dilemma was around 1 minute long.

There were two passengers in the vehicle each time, introduced as a female friend and a friend's son. The participant was in the driver's seat. Participants learned during pre-training, that the vehicle could not harm human pedestrians outside the vehicle and would automatically brake if there was a person in the path of the vehicle, even if that risked the lives of people inside the car. Participants were aware that they could 'override' the auto-braking function in an emergency situation, which would keep the car moving forward, causing fatal harm to any pedestrians in the path of the vehicle. The override button was

either a green button in the middle of a steering wheel, or a foot-pedal on the floor. Pressing 'override' triggered a high-pitched beep sound. The car AI communicated warnings about pedestrians and collisions with other vehicles via audio and a heads-up-display system, along with some other information such as a welcome message, or road conditions. In each situation, the first warning (~20 seconds in) directed attention towards a hazard such as a lorry on fire (Dilemma 1), or a motorcyclist that had fallen in the path of the vehicle (Dilemmas 2/3). The second warning was the collision alert whereby the AI informed participants if the car braked to avoid a pedestrian, there would be a collision with another vehicle coming towards it which was presented ~15 seconds after warning 1 in Dilemma 1, and ~5 seconds after warning 1 in Dilemma 2 and 3. In all dilemmas if the participants omitted to act, the car would automatically brake, and a collision would ensue. The seriousness of the collision was communicated quickly by audio-visual 'risk red' and 'risk yellow' warnings emphasised in the pre-training, which served as the 'incongruent' and 'congruent' conditions respectively. 'Risk red' indicated fatal harm to people inside the vehicle if the car remained stationary. 'Risk yellow' indicated damage to the vehicle or minor injuries to passengers if the car remained stationary. The vehicles in each dilemma were appropriate to the 'risk' of fatality (Table 1). Following the collision warning, participants were told that they could press the 'override' button to keep the car moving and reminded of the outcome of pressing the override button. The pedestrian was either crossing the road in front of the vehicle and screamed (1) or was on the road and in distress after falling off a motorbike (2/3). In Dilemma 1 (incongruent), the override button was only activated when the audio collision warning was finished to ensure the car was a certain distance away from the pedestrian, which was discussed in pre-training. If 'override' was pressed before it was activated, participants were informed by the experimenter when the button became activated (also discussed during training). Participants had ~15 seconds to make a decision from the onset of the collision warning, and ~5 seconds from the end of the collision warning. No graphic or gruesome information was presented to participants. The scene ended just before hitting the pedestrian or just before the collision with the other vehicle. A text display after each dilemma informed participants of the outcome of their decision.

Table 1.

Risk red and risk yellow dilemmas

	Risk red (incongruent)	**Risk yellow (congruent)**
Fatalities to all passengers if 'override' not pressed?	Yes	No
Vehicles involved in collision in each matched dilemma pair	Dilemma 1: Lorry Dilemma 2: Bus Dilemma 3: Train	Dilemma 1: Campervan Dilemma 2: Car Dilemma 3: Tram

Foot-pedal and button-press

The foot-pedal and steering wheel button are presented in Figure 2. The action of pressing was recorded with a mouse button beneath the surface of the green button and foot pedal. Actions were carried out with either the right hand (button) or right foot (pedal). Participants practiced these actions prior to the task and asked to place their foot at the base of the foot-pedal, or close to the button on the steering wheel so they could make the action without difficulty. Participants were also asked to have their right hand on the steering wheel in the foot pedal condition.

Figure 2.

Button-press and foot-pedal devices

Measures of moral behaviour

As Conway and Gawronski's (2013) process dissociation formula for calculating harm-aversion (deontological parameter) and outcome-maximisation (utilitarian parameter) scores requires no zero values for the outcome-maximisation parameter (which occurs when no non-utilitarian responses are made), we were unable to establish harm-aversion scores for many participants as the majority of people choose to carry out harmful action on all incongruent trials. A greater number of trial-pairs would potentially reduce the likelihood of this error and increase variation in responses. Subsequently, our measures of moral behaviour included: 1) percentage of harmful acts carried out on incongruent trials (where harming the pedestrian would save the lives of all people in the vehicle), 2) percentage of harmful acts carried out on congruent trials (where harming the pedestrian did not save the lives of people in the vehicle), 3) response time for harmful action on incongruent trials.

Procedure

Participants were asked to not consume any food or drink for at least 2 hours or consume any alcohol or (unprescribed) drugs 24 hours prior to the study. After providing informed consent, participants completed the anxiety and interoception questionnaires at a desktop computer and provided basic demographic information. Height and weight were measured, and the heartbeat counting task completed. Twelve participants who completed the heartbeat detection and water load tasks in a previous study, did not complete these tasks as they provided informed consent for their data to be re-used. Physiological data recording equipment was setup and participants read through a training sheet on the computer to explain the VR environment and their role in the task. Participants were expressly reminded they could end the VR experiment at any time by saying a safe-word or removing the headset. The 'override' button was indicated as either the foot pedal or button on the steering wheel. They were unaware of the other action-type. The experimenter asked several questions to assess they understood the procedure. Participants first experienced an orientation scene in VR where they were sitting in the moving car with other passengers, the headset was then removed. Physiological data acquisition began 2-3 minutes before the moral dilemma task while participants were seated at the computer. Participants were asked to keep as still as possible. Blood pressure was taken twice before the experiment. Their left arm remained raised and resting on a table in front of them (palm-up) throughout the task. They were also

asked to place their right hand on the steering wheel. The VR headset was placed again, and the VR moral dilemma task began, starting with a 20 second rest screen. They experienced 6 moral dilemmas in the same order incongruent (1); congruent (2); incongruent (3); congruent (1); incongruent (2); congruent (3). In each dilemma, following their choice a text screen stated the outcome, followed by a 20 second rest screen. Participants were in the VR environment for about 10 minutes in total. Immediately following the task blood pressure was measured. There was a short rest period before participants completed the presence and motion sickness questionnaires. Finally, they completed the water-load task and hunger/thirst scales. The experiment took 1-1.5 hours to complete. Participants were debriefed and thanked for their time.

Physiological data collection

Impedance cardiographic (ICG), electrocardiographic (ECG) and electrodermal signals were recorded using a wireless BIOPAC MP160 (BIOPAC Systems, 2020). Cardiac output (CO; blood volume ejected from the heart at systole), systemic vascular resistance (SVR; resistance in circulatory system to achieve blood flow), pre-ejection period (PEP; latency between depolarisation of the left ventricle of the heart and ventricular ejection), heart rate (HR; beats-per-minute), and skin conductance activity (SC) were measured. Physiological data were digitised, stored and analysed using *Acqknowledge* software (BIOPAC Systems, 2015). To capture ICG, electrodes were placed either side of participants neck, and upper torso. An Einthoven lead II configuration was used to collect ECG data. Electrodes on the index and second finger captured electrodermal data. Wireless ICG, ECG and SC amplifiers were attached with a chest-strap and wrist-strap. *Acknowledge* software was used to calculate SVR, PEP, CO and HR. A clinically validated blood pressure monitor (OMRON) was used to calculate mean arterial pressure (MAP) (Cushman et al., 2012). Blood pressure was measured twice before entering the VR environment and once immediately after to calculate an average MAP for the ICG analysis, and was calculated using the formula:

$$MAP = (systolic\ pressure + 2(diastolic\ pressure))/3$$

Physiological data processing

A 12.5msec delay was added to digital event marker channels before data processing to correct signal latency from the wireless amplifiers. Waveforms were checked and cleaned for motion-artefacts and noise. SVR, CO and PEP was calculated for each participant in *Acqknowledge* (BIOPAC Systems, 2015). Digital event markers were recorded on the data file to ensure exact measurement-windows for analysis. Programming scripts were used to calculate mean CO, SVR, HR, PEP and SC data for the three measurement windows: 1) start to the second collision warning; 2) second collision warning to action-choice; 3) action choice to end of rest period. A 1-second offset was applied to SC measurement windows to account for signal delay. Importantly, measurement window 3) for the final rest period (congruent 3) was shorter than the other dilemmas, as participants removed the headset shortly after presentation of the final text screen.

Cardiac reactivity

All raw physiological data was transformed into relative values (*z*-scores) within each participant's dataset which can improve reliability and validity of individual differences analyses (Boucsein et al., 2012; Braithwaite et al., 2013; Braithwaite & Watson, 2015). Means and standard deviations used to compute *z*-scores were calculated within measurement windows 1-3. CO and SVR scores were transformed into one threat-challenge reactivity score (see Scheepers et al., 2012; Turner et al., 2013). According to the biopsychosocial model (Blascovich & Mendes, 2010; Mendes et al., 2007), cardiovascular states are a manifestation of an evaluation of the demands of a situation. A challenge motivational state occurs when resources meet or exceed demands leading to an efficient cardiovascular response: an increase in CO and reduction in SVR. When demands outweigh resources a threat motivational state arises: a decrease in CO and increase in SVR. To calculate the threat-challenge reactivity index, raw SVR scores were subtracted from CO scores. Higher scores reflect a higher challenge state, and lower scores indicate a higher threat state. Shorter PEP is associated with stronger myocardial contractility and sympathetic autonomic activation on active coping tasks (Sherwood et al., 1990). Threat-challenge reactivity (TCR), heart rate (HR) and pre-ejection period (PEP) values for all trials within each measurement window (1-3) were transformed into *z*-scores and then *t*-scores to provide a relative measure of HR and TCR within each timeframe on congruent and incongruent trials.

Skin conductance

Skin conductance data was transformed into Area Under the Curve (AUC; Naqvi & Bechara, 2006). A moving-average smoothing filter was first applied to the SC waveform (2000 samples). The waveform was resampled at 125Hz, and a difference transformation (interval = 6 samples) was applied. Area under the curve represents the area bounded by the curve and the chord connecting the curve with endpoints of each measurement window (Naqvi & Bechara, 2006). Area was divided by the time in seconds for that measurement interval, to calculate a rate of micro-siemens per second (AUC). This method provides an indication of amplitude and temporal features of electrodermal activity and removes the need for subjective interpretation about the presence of SC responses (Naqvi & Bechara, 2006). A '*log (AUC + 1)*' transformation was applied to raw AUC scores within each measurement window to allow for zero values and normalise skew (Braithwaite et al., 2013; Braithwaite & Watson, 2015). Within each measurement window, logged AUC scores were transformed into *z*-scores, and finally *t*-scores to remove negative values (see Braithwaite & Watson, 2015). Means and standard deviations used for *z*-score calculations were based on all congruent and incongruent within each measurement window. Average *t*-values for measurement windows 1-3 were calculated for the incongruent and congruent trials.

Results

Statistical analysis was carried out using SPSS v26. A bootstrapping method (x1000) was applied to confidence intervals for ANOVA and regression analyses. A Shapiro Wilks test confirmed normal distribution of the physiological variables *(p*>.05) except for heart rate in the pre-decision timeframe and electrodermal activity in the post-decision timeframe. Levene's tests confirmed equality of variances of the physiological variables between foot-pedal and button-press conditions (*p*>.05). Three outliers were identified for the AUC data in the pre-warning timeframe but retained in the interests of sample size. Data are mean ± standard deviation unless stated otherwise.

Correlations

We explored bivariate correlations of potential confounding variables of age, sex, and anxiety with the dependent variables of harm acceptance on congruent and incongruent trials, and response time on incongruent trials. Age, sex and anxiety did not predict choices to harm

on congruent or incongruent trials or response time to harm on incongruent trials. We further explored any confounding influence of age, sex, BMI and anxiety with any of the interoception variables, and any relationships between interoception variables (Table 2). Age, sex and BMI did not predict any of the interoception variables. Surprisingly, state anxiety was negatively correlated with interoceptive sensibility (p=.015), suggesting that people with a greater tendency to pay attention to internal sensations were less anxious at the time. None of the interoception variables showed correlation with each other, supporting the independence of these dimensions (Garfinkel et al., 2015).

Table 2.

Bivariate correlations between potential confounding variables, key dependent variables and interoception

		2	3	4	5	6	7	8	9	10	11	12
1.	State anx	.675**	.004	-.089	.094	-.375	-.165	-.018	-.092	-.480*	-.090	-.385
2.	Trait anx	1	-.011	-.083	-.283	-.124	.136	-.181	.038	-.196	-.078	-.190
3.	Sex		1	.022	.203	-.019	.171	.167	-.170	.269	.102	-.067
4.	Age			1	.184	-.114	-.212	.022	.187	-.008	-.259	-.081
5.	RT (inc)				1	-.149	-.764**	.206	-.056	.077	.112	-.041
6.	Gastric int.					1	.047	.229	.153	.159	-.025	.089
7.	Harm (Inc)						1	.055	.072	.102	-.035	-.056
8.	Harm (Con)							1	-.130	.260	-.119	-.240
9.	BMI								1	.103	-.320	-.004
10.	IS									1	.290	-.044
11.	IAc										1	.213
12.	IAc-meta											1

Note: 1. State anxiety, 2. Trait anxiety, 3. Sex (male=1), 4. Age, 5. Response time for harmful action, 6. Gastric interoception, 7. Harmful action incongruent trials, 8.Harmful action congruent trials, 9. Body mass index, 10. Interoceptive sensibility, 11. Interoceptive accuracy, 12. Interoceptive meta-cognitive awareness

H1. Effect of dilemma, action-type and timeframe on physiological arousal

We investigated whether physiological arousal was affected by dilemma type, timeframe and action type. None of the physiological variables were correlated except HR

and threat-challenge reactivity, which were positively correlated within each timeframe (p<.05). For each physiological arousal index, we conducted 2 (dilemma-type: congruent vs. incongruent) x 3 (timeframe: pre-warning, pre-decision, post-decision) mixed-factor ANOVAs. Action-type (foot-pedal vs button-press) was entered as a between participant's factor. Means and standard deviations for each condition are in Table 3. To account for pairwise comparisons p-values (α) were set to <.025. We did not find main effects of dilemma-type, action-type or time period for HR, threat-reactivity or PEP (p>.05). For AUC, there was a main effect of dilemma type, $F(1, 23) = 9.791$, p=.005, partial η2=.299, and a significant interaction between dilemma-type and timeframe for people in the button-press condition, $F(2, 36)=6.164$, p=.005, partial η2=.255, but not for the foot-pedal condition, $F(2, 10)=1.374$, p=.297, partial η2=.216. Specifically, AUC was higher post-decision (54.65 ± 3.75) compared to pre-decision (51.55 ± 4.33) in the incongruent conditions, but lower post-decision (45.34 ± 3.75) compared to pre-decision (48.84 ± 4.33) in the congruent conditions, which was significant, p=.006, 95% CI [-6.065 to -.940]. We also found a main effect of dilemma type on AUC for the foot-pedal condition in the pre-decision timeframe, $F(1, 5) = 11.05$, p=.021, partial η2=.688. Specifically, AUC was significantly higher in the incongruent condition (54.99 ± 3.68) compared to the congruent condition (45.00 ± 3.68) prior to carrying out harm in the foot-pedal condition, providing partial support for H1.

Table 3.

Physiological responses (HR, PEP, AUC, Threat-challenge reactivity) for button-press and foot-pedal conditions for incongruent trials

		Button-press			**Foot-pedal**		
		Pre-warning	Pre-decision	Post-decision	Pre-warning	Pre-decision	Post-decision
Threat-Challenge	*Mean*	50.45	52.05	49.59	49.99	49.51	50.78
	SD	4.03	5.08	4.43	3.59	3.65	6.01
AUC	*Mean*	52.33	51.15	54.65	52.10	54.99	53.07
	SD	3.13	4.33	3.75	3.61	3.68	4.39
HR	*Mean*	52.06	51.33	50.21	50.69	51.23	52.02
	SD	3.24	4.85	4.42	4.02	5.29	5.67
PEP	*Mean*	49.72	49.29	48.61	50.35	49.15	49.95
	SD	3.05	4.46	3.50	5.02	5.41	2.74

Note. *Mean t scores > 50 indicate increased HR, TCR, and AUC relative to congruent trials*

H2. Physiological arousal, action-type and behaviour

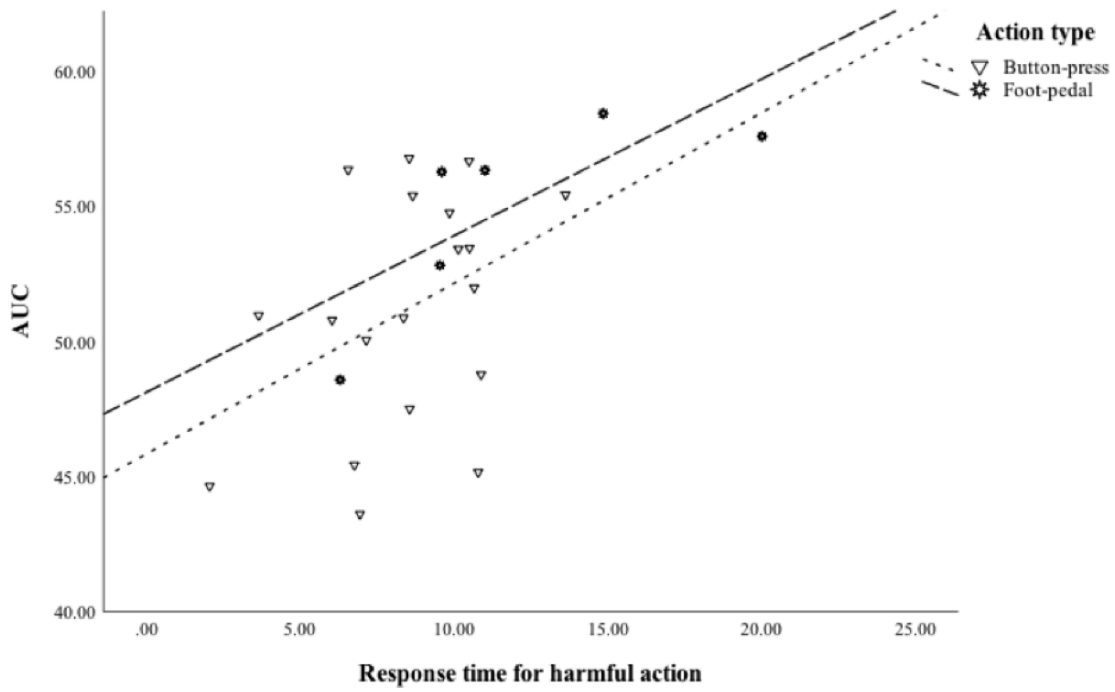
We tested whether physiological arousal and an interaction between group type and physiological arousal predicted behaviour in the moral dilemma task. Multinomial logistic regressions were conducted for the dependent variables 1) frequency of harmful action on incongruent trials, 2) frequency of harmful action on congruent trials. A hierarchical regression was carried out for the dependent variable 3) response time for harmful action on incongruent trials. One person was excluded from RT analysis in the foot-pedal condition as

they did not carry out any harmful actions. Twenty-one people carried out harm on 3/3 incongruent trials, two people carried out harmful action on 2/3 trials, and one person on 1/3 trials. Interaction terms for action-type*physiological index for the pre-decision timeframe on incongruent trials were entered as predictors (e.g. AUC*Foot-pedal).

There was no interaction between action-type and physiological arousal for predicting frequency of harmful action on incongruent trials, X^2 (12) =15.958, p=.193, or on congruent trials, X^2 (4) =4.730, p=.316. Although the likelihood ratio for action-type*HR approached significance, X^2 (1) =3.326, p=.068, suggesting people who carried out harmful action on congruent trials in the foot-pedal condition had relatively higher HR on these trials, compared to the button-press condition. For response time, physiological indexes for the pre-decision timeframe were entered at step 1, and the interaction terms (e.g. action-type*AUC), at step 2. Model 1 was significant, R^2=.384, F (4, 19) = 2.961, p=.047. Specifically, greater AUC significantly predicted longer response times (b=.588, p=.005). Contrary to H2, the addition of the interaction terms in model 2, did not significantly improve model fit, R^2=.535, F (4, 15) = 2.153, p=.095, suggesting action-type did not influence this effect. Figure 3 shows a positive relationship between AUC and RT in both the foot-pedal and button-press conditions.

Figure 3.

Scatterplot with fit line for AUC and response time in button press and foot-pedal conditions



H3. Interoceptive accuracy and sympathetic reactivity

We explored whether sympathetic reactivity in the incongruent conditions relative to congruent conditions was associated with interoceptive accuracy (IAc). Using a median-split procedure we calculated high and low reactivity groups for the physiological variables in the pre-decision time frame. High reactivity was defined as greater sympathetic activation in incongruent trials characterised as an increase in HR and AUC, and a decrease in PEP and threat-challenge reactivity (relative to congruent trials) (Table 4). This method of comparing interoceptive accuracy among high and low reactivity groups was chosen to compare our findings with a recent study using several of the same physiological parameters, comparing interoceptive accuracy differences for participants high and low in cardiovascular reactivity during an exercise task (see Pollatos et al., 2007a). As participants' physiological data was standardised within-subject, this provided a comparable measure of physiological 'reactivity'

for each person, that accounted for inter-individual differences in baseline physiological reactivity.

IAc was the dependent variable, and PEP, HR, Threat-challenge reactivity and AUC reactivity groups were entered as between subjects' factors in a 4-way ANOVA (Table 4). In support of H3, IAc was significantly greater in the low PEP group, F (1, 14) =11.577, p=.004, partial η2=.453, but not significantly different between groups for HR, F (1, 14) =1.935, p=.186, partial η2=.121, threat-challenge reactivity, F (1, 14) =.186, p=.673, partial η2=.013, or AUC, F (1, 14) =.196, p=.664, partial η2=.014. Interestingly, IAc showed a positive relationship with PEP in the pre-decision timeframe on incongruent trials (r=.449, p=.024) (Figure 4). Greater PEP is associated with reduced sympathetic autonomic arousal (Sherwood et al., 1990), and motivational challenge states (Tomaka et al., 1993) and may indicate a greater psychological challenge state among higher IAc perceivers on trials where moral conflict was present (incongruent), compared to trials where there was no moral conflict.

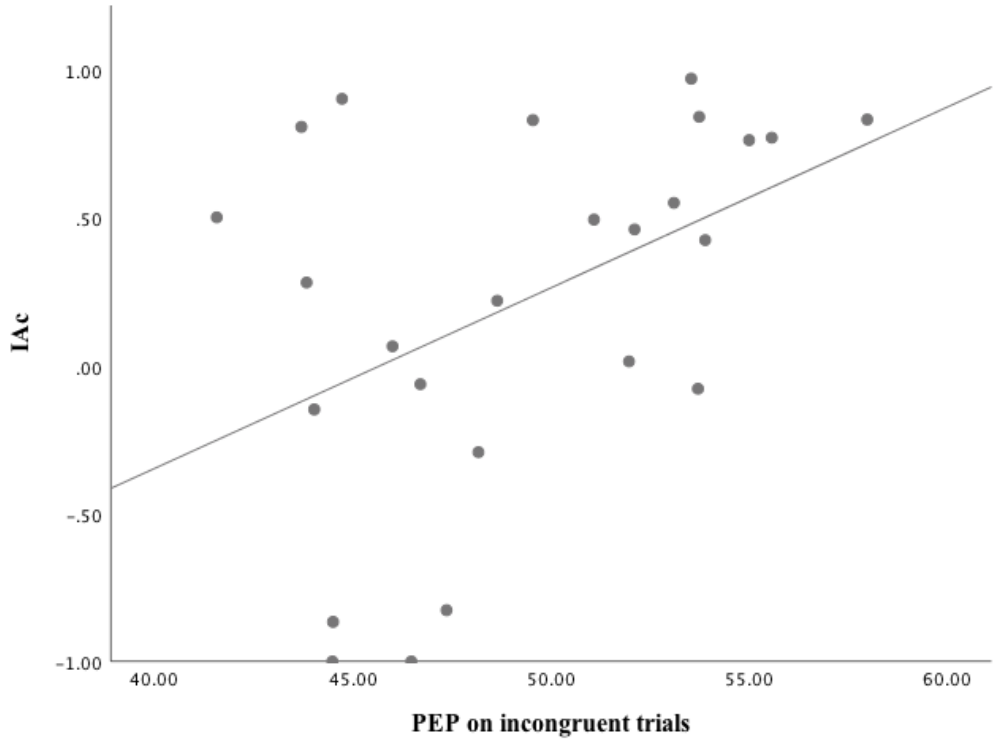
Table 4.

Means and standard deviations for Interoceptive Accuracy (IAc) for harmful action in high and low sympathetic reactivity groups

		High PEP	Low PEP	High HR	Low HR	High threat-reactivity	Low threat-reactivity	High AUC	Low AUC
IAc	*Mean*	-.135	.548	.042	.413	.360	.068	.325	.107
	SD	.683	.333	.653	.557	.562	.679	.628	.629

Figure 4.

Scatterplot showing relationship between Interoceptive accuracy (IAc) and pre-ejection period (PEP) on incongruent trials



H4. Interoception, presence and virtual reality sickness

We explored whether interoception influenced experiences of presence in the VR environment. We carried out a hierarchical linear regression model, inputting the between subjects' factor action-type and trait and state anxiety scores at step 1, as potential confounding variables (Bouchard et al., 2008). The interoception variables (IAc, IS, IAc-meta and gastric interoception) were entered at step 2. Model 1 was not significant, R^2=.044, F (3, 21) = .323, p=.809. Model 2 was also not significant, R^2=.410, F (4, 17) = 1.686, p=.179, but a significant coefficient for IAc indicated that interoceptive accuracy was positively associated with greater experiences of presence b=.487, p=.031, providing support for H4. However, when using the conservative SUS approach that calculates presence as the number of 6 or 7 scores provided, IAc did not significantly predict presence in a multinomial regression, x^2 (4) = 2.60, p=.627. Therefore, IAc may be associated with greater experiences of presence when calculated as a continuous SUS score but not when conceptualised as the

presence of high scores. This is potentially due to the small sample size, with 52% of people scoring zero using the 6 or 7 scoring approach. We did not find any of the interoception variables significantly predicted motion-sickness in separate linear regression models (p>.05), suggesting interoceptive sensitivity did not influence oculomotor or disorientation disturbances associated with the VR environment.

H5. Presence and physiological arousal

To test whether experiences of presence were associated with greater physiological arousal on incongruent trials, we carried out a hierarchical linear regression model inputting presence as a continuous dependent variable. Anxiety, action-type and IAc were entered at step-1. The high and low sympathetic reactivity groups for PEP, HR, AUC and threat-challenge index in the pre-decision timeframe on incongruent trials were entered at step-2. Model 1 was not significant, R^2=.270, F (4, 20) =1.854, p=.158, but the coefficient for IAc was significant, b=.479, p=.022. The addition of the physiological variables at step-2 did not significantly improve model fit, R^2=.494, F (4, 16) = 1.956, p=.121. However, the coefficient for IAc remained significant, b=.522, p=.035, suggesting that IAc independently predicted experiences of presence when accounting for physiological reactivity on the incongruent trials. Additionally, the coefficient for threat-challenge reactivity group approached significance (b=.595, p=.067). People in the 'low' threat-challenge reactivity group demonstrated a relatively greater psychophysiological 'challenge' state [as opposed to 'threat' state] on incongruent trials and experienced greater levels of presence (mean=3.54, SD=1.07) than people in the high threat-challenge reactivity group (mean=2.97, SD=1.09).

H6.1. PEP reactivity, action-type and response time

To investigate an interaction between PEP reactivity and action-type on response time, we conducted a 2 (action-type: button-press vs foot-pedal) x 2 (PEP reactivity: high vs low) ANOVA. There was no main effect of action-type, F (1, 20) =2.547, p=.126, partial η2=.113, or PEP reactivity, F (1, 20) =.240, p=.629, partial η2=.012, but the interaction between action-type and PEP reactivity approached significance, F (1, 20) =3.421, p=.079, partial η2=.146. We conducted planned pairwise comparisons (Faraway, 2015) with Bonferroni adjustment and found an effect of action-type in the high PEP reactivity group,

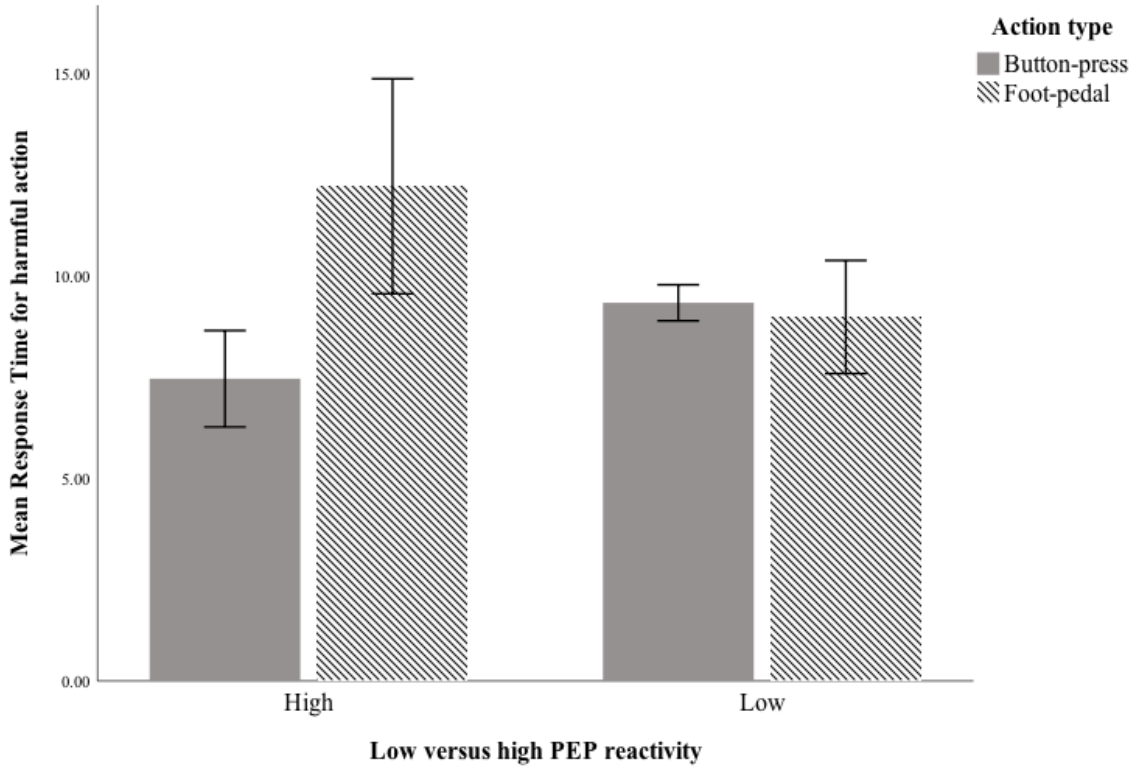
$F(1, 20) = 5.072$, $p = .036$, partial $\eta 2 = .202$. Specifically, the foot-pedal condition was associated with longer reaction times for carrying out harmful action, but only for people demonstrating higher [sympathetic] PEP reactivity on incongruent trials relative to congruent trials (Figure 5).

H6.2. Interoceptive accuracy, action-type and response time

Following the findings in H6.1 and the association between PEP and interoceptive accuracy (H3), we explored the possibility that action-type moderated the relationship between IAc and response time for harmful action (RT). None of the interoception measures (IAc, IS, IAc-meta and gastric interoception) were significantly associated with response time in a multiple regression model, $R^2 = .156$, F (4, 19) = .879, $p = .495$. IAc was also not significantly different between the button-press and foot-pedal conditions, t (22) =.051, $p = .960$. A hierarchical regression model was used to test the interaction between IAc and action-type on RT. IAc and dummy variables for action-type were entered at step 1, and the interaction terms for action type and IAc entered at step 2. Model 1 was not significant, $R^2 = .078$, F (2, 21) =.882, $p = .429$, but the addition of the IAc*action-type at step 2 led to a statistically significant increase in variance explained, R^2 chng=.191, F chng= 5.238, *Sig*. F chng =.033 in the model, $R^2 = .269$, F (1, 20) =2.453, $p = .093$. Simple slopes analysis showed there was a statistically significant positive relationship between IAc and response time in the button-press condition ($b = .934$) but a significantly negative relationship between IAc and response time in the foot-pedal condition ($b = -.517$), $p = .033$, 95% CI [.427 to 9.211]. These findings suggest that IAc predicted shorter response times in the foot-pedal condition, and longer response times in the button-press condition (Figure 6).

Figure 5.

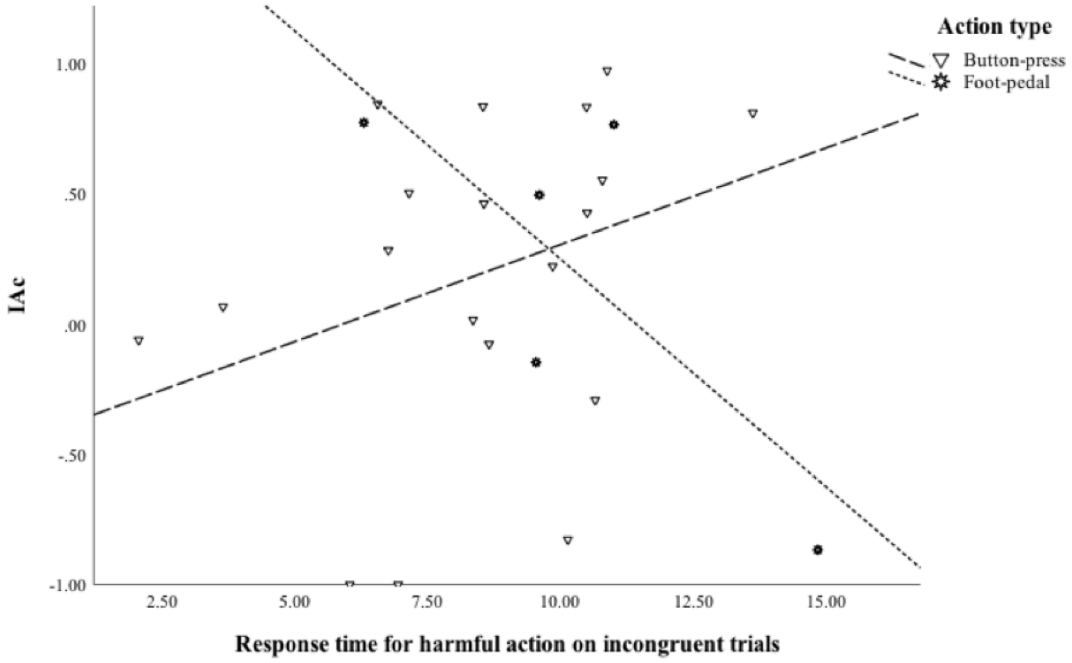
Bar chart showing response time for high and low PEP sympathetic reactivity groups in button-press and foot-pedal conditions



Note. Error bars are +/- 1 standard error

Figure 6.

Scatterplot with fit lines for relationship between interoception accuracy and response time for foot-pedal and button-press conditions



Discussion

We investigated whether physiological arousal, interoception and action-type influenced harmful actions in moral dilemma driving tasks. The majority of participants carried out harmful action on incongruent trials (i.e. where harm maximised outcomes). Very few people chose to carry out harm on congruent trials, indicating that people were typically averse to carrying out harm that did not maximise outcomes. A high level of utilitarian responses is consistent with previous VR studies, for example, Francis et al (2017) found >56% of people chose to carry out utilitarian action in a VR Footbridge dilemma, despite only 10% of people condoning this response when asked to provide a moral judgment of the text-based dilemma. As Patil et al (2014) proposed, in VR, the contextual saliency of the potential negative outcomes of not carrying out harmful action can overcome the potential negative value associated with carrying out a harmful act. Electrodermal activity was higher in incongruent trials and the strength of electrodermal activity predicted longer response

times for harmful action but showed different patterns of temporal activity for the button-press and foot-pedal conditions. Overall, pre-ejection (PEP) period shortened in the pre-decision time frame relative to the pre-warning timeframe, but this reduction was greater for the foot-pedal condition, although not significant. Better heartbeat perception (IAc) predicted a lengthening of pre-ejection period prior to carry out harmful action which suggests a reduction in sympathetic autonomic arousal and greater psychophysiological challenge states. Sympathetic PEP reactivity was associated with longer response times in the foot-pedal condition, and shorter response times in the button-press condition. Relatedly, we found the relationship between IAc and response time was moderated by action-type. IAc was also associated with greater experiences of presence. These findings are discussed further below.

In H1, we explored the effects of dilemma type, timeframe and action-type for each of the physiological variables; heart rate (HR), electrodermal activity, threat-challenge reactivity and pre-ejection period (PEP). Electrodermal activity was highest in the *pre*-decision timeframe in the foot-pedal condition but highest *post*-decision in the button-press condition. Greater anticipatory autonomic arousal prior to pushing the foot-pedal may be due to a greater aversive response to the pedal action than the button-press, where arousal was higher after the action was taken. Previous work using text-based dilemmas has found stronger electrodermal responses predict harm-rejection judgments, especially when harm requires direct interpersonal force (e.g. McDonald et al., 2017; Moretto et al., 2010). However heightened electrodermal activity [relative to text-based dilemmas], has been associated with carrying out 'utilitarian' harmful actions in virtual reality dilemmas (e.g. Patil et al., 2014). This is consistent with our finding that people experienced greater electrodermal responses on trials where harmful action was required to achieve a utilitarian outcome, but harmful action was still carried out.

We did not find any significant effects of dilemma type, timeframe or action-type for indices of HR, threat-reactivity or PEP. Prior work has found that increased systemic vascular resistance (SVR; Cushman et al., 2012) and shorter pre-ejection period is associated with an aversion to action-based harm (Parton & McGinley, 2019). A larger sample size may be more effective at detecting these effects. Importantly, a shortening of PEP observed on incongruent trials (see Table 3) in the pre-decision timeframe (relative to congruent trials) is likely partially due to the exertion required to carry out the actions (pushing pedal/button) (Sherwood et al., 1990) compared to congruent trials where action was typically not performed.

In partial support of H2, we found electrodermal activity significantly predicted longer response times for carrying out harmful action, but not the choice to carry out harmful action or not. The measure of electrodermal activity used (AUC; Naqvi & Bechara, 2006), provides a measure of both the temporality and amplitude of emotional arousal and suggests longer response times were associated with a greater emotional response to the prospect of carrying out harm. Other work has found longer response times when contemplating personal as opposed to impersonal dilemmas (Moretto et al., 2010) and a recent study found greater electrodermal responses predicted longer reaction times on a text-based moral judgment task (Brown, Proulx, & Fraser, 2020b). Electrodermal activity has also been associated with cognitive load in driving simulations (Solovey et al., 2014), which may have been associated with lengthier deliberation of the costs and benefits of harmful action in this study, before making a choice. Contrary to H2, there was no interaction between electrodermal activity and action-type, suggesting the influence of electrodermal activity on response time was independent of the typicality of the action.

In H3 we found high and low PEP reactivity in the pre-decision timeframe on incongruent trials predicted interoceptive accuracy (IAc), but not as expected. Whereas prior research has found heartbeat detection ability predicts shorter PEP reactivity to aversive stimuli, arithmetic tasks and physical stressors (Eichler et al., 1987; Eichler & Katkin, 1994; Pollatos et al., 2007b) we found IAc was associated with longer PEP on incongruent trials. Shorter PEP has been associated with active coping on motor tasks (Sherwood et al., 1990) and carrying out simulated harmful acts requiring physical effort (Parton & McGinley, 2019). Therefore, people with better heartbeat detection ability showed reduced sympathetic activation prior to carrying out harmful acts, compared to people with poorer heartbeat detection ability. Interestingly, in active coping motor-tasks people showing greater reactivity in systemic vascular resistance demonstrate a lengthening of PEP, whereas people who show greater cardiac output reactivity show a shortening of PEP (Sherwood et al., 1990). This suggests that individual differences in cardiovascular reactivity may influence the direction of PEP when faced with active coping tasks. Furthermore, on arithmetic active coping tasks, longer PEP has been associated with psychological challenge states, and shorter PEP with psychological threat states (Tomaka et al., 1993). It is possible that people scoring highly on IAc, appraised the task of carrying out harmful action as more of a challenge than a threat.

In line with the above explanation, there is evidence to suggest IAc may contribute to more effective emotional regulation of aversive states (Füstös et al., 2013), including the reappraisal and suppression of emotional responses (Kever et al., 2015). IAc has also been associated with more pronounced heart rate modulation in response to emotional stimuli (Pollatos et al., 2007b). It has been suggested that higher IAc may facilitate a more flexible use of emotional regulation strategies (Kever et al., 2015), which could support more adaptive behavioural responses (Aldao, 2013) to environmental stimuli, especially when there is little time to act. In a related study, Brown et al (2020b), found that IAc moderated the influence of an aversive physiological response on harm-aversive judgments, indicating better heartbeat detectors may be less reactive to negative emotional states generated by the prospect of harm, that may otherwise lead people to reject harmful action (e.g. Cushman et al., 2012; Greene et al., 2001; Parton & McGinley, 2019). As Farb et al (2015) suggest, interoception is a dynamic process, involving the integration of interoceptive sensations with appraisals of interoceptive states to facilitate the selection of an adaptive behavioural response. It is possible that people's ability to detect cardiovascular signals with precision, may support more adaptive modulation of cardiovascular arousal in scenarios where aversive harmful action is required to achieve a utilitarian outcome. Alternatively, or perhaps relatedly, IAc may have predicted reduced effort exerted when carrying out harmful actions during the task. People with higher IAc ability have shown to exert less physical and cardiovascular effort during exercise, potentially due to their enhanced ability to perceive bodily signals associated with physical and cardiovascular load (Herbert et al., 2007). It may be that these people were better able to regulate their physical and cardiovascular exertion when performing the actions. If so, this may undermine future research efforts to investigate associations between cardiovascular parameters of action-based harm aversion and moral behaviour.

In H6.1, we found that higher sympathetic PEP reactivity on incongruent trials was associated with longer reaction times to carry out harmful action, but only for the foot-pedal condition. For the low PEP reactivity group, response time was similar for the button-press and foot-pedal conditions. Relatedly, in H6.2, we found interoceptive accuracy (IAc) predicted longer response times in the button press-condition, but shorter response times in the foot-pedal condition. This speaks to the finding in H3, whereby high PEP reactivity on incongruent trials in the foot-pedal condition was associated with the longest reaction times. IAc however, predicted lower [sympathetic] PEP reactivity on incongruent trials, and was

therefore associated with quicker response times for harmful action in the foot-pedal condition. Afferent heartbeat signals have shown to influence the processing of fearful stimuli (Garfinkel & Critchley, 2016), and whereas some indices of cardiac activity such as heart rate deceleration have shown to support richer sensory processing, others such as heart rate acceleration, are associated with shallower sensory processing to support action and quicker response times (Lacey & Lacey, 1978). IAc may facilitate faster response times due to low sympathetic PEP reactivity if afferent signals associated with the lengthening of PEP inhibits richer sensory processing; but in the absence of empirical support this supposition remains speculative.

More generally, when faced with aversive emotional stimuli, IAc may facilitate superior cardiovascular modulation (Pollatos et al., 2007b) and emotional regulation strategies (Füstös et al., 2013; Kever et al., 2015) associated with the aversive action, that potentially allow them to engage in this action more quickly to achieve a utilitarian outcome. The positive relationship between IAc and response time in the button-press condition, however, challenges this explanation. This may be because, in the button-press condition, people in the high PEP reactivity group demonstrated the fastest response times overall, and comparatively, higher IAc people may have provided slightly longer response times. Importantly, our sample size was underpowered to conduct moderation analyses and should be considered cautiously.

IAc was also associated with greater experiences of presence (H4) even when accounting for high and low physiological reactivity on incongruent trials (H5), but only when assessed as a continuous scale and not the presence of high scores (Slater et al., 2000). Interestingly, people demonstrating physiological ‘challenge’ states on incongruent trials (captured by the threat-challenge reactivity index), experienced greater levels of presence in VR. Consistent with this, psychological ‘flow’ states (characterised by complete immersion in what one is doing; Nakamura & Csikszentmihalyi, 2005) is typically associated with increased cardiac output (Ullén et al., 2013), which has also been linked with greater presence (Weibel et al., 2008). Diemer and colleagues (2015) suggest people partly determine their sense of cognitive presence based on their level of emotional arousal. Although some work has not found associations between emotional experience and presence (Felnhofer et al., 2015), a recent study (Heeter et al., 2020) found greater self-reported attention towards internal sensations predicted greater experiences of presence, using the Multidimensional

Assessment of Interoceptive Awareness scale (MAIA; Mehling et al., 2018). Seth et al's (2012) predictive coding model of conscious presence holds that interoceptive or emotional states [in VR] are continuously compared with expected states, and that presence arises when the incongruence between actual and expected signals is resolved. Farb et al (2015) suggest that in instances of perceptual inference, the greater precision weighting of current sensations over rigid expectations ('priors'), may support experiences of embodied presence which could explain the positive association between heartbeat detection ability and presence in the current study.

Limitations

Our sample was smaller than planned and group conditions unequal in size due to the onset of coronavirus and is underpowered to detect some of the effects tested. Likewise, the significant findings observed should be considered prudently, as further higher-powered replications are required to determine the robustness of these effects. For example, we chose not to conduct a MANOVA including all arousal indexes in H1, due to the substantial reduction in degrees of freedom in smaller samples. It is worth noting that for the interaction between electrodermal activity, dilemma type and timeframe (H1), effect sizes for the two conditions (button-press and foot-pedal) were of similar size, which suggests that an increase in sample size could be promising to detect physiological effects of action-type in future iterations using this paradigm.

Secondly, the actions required to carry out harm in the foot-pedal and button press conditions aimed to be as close to in task-involvement and physical labour as possible but were qualitatively different, i.e. a pushing action with right hand versus pushing action with right foot. Therefore, we cannot rule out that these actions contributed to differential effects on the cardiovascular parameters measured (Parton & McGinley, 2019; Sherwood et al., 1990). Future iterations could contrast a greater range of physically 'matched' typical and non-typical driving actions, to observe differences in their effect on cardiovascular reactivity. Furthermore, despite piloting the relative 'aversiveness' of the foot-pedal and button-press with a small audience, we did not capture participant reports of the relative aversiveness of the food-pedal versus button-press which limits our understanding of whether the foot-pedal action was subjectively experienced as more aversive.

In addition, it was not possible to calculate harm-aversion and outcome-maximisation parameters as outlined in Conway and Gawronski's (2013) text-based method. A greater

number of trials, situations, or a higher level of uncertainty about the outcome of carrying out harm may increase variation of responses to calculate these parameters. Moreover, the heartbeat detection method (Schandry, 1981) used to calculate interoceptive accuracy has been criticised for its validity, as some studies have shown that heartbeat detection is influenced by other factors, such as subjective beliefs about resting heart rate (Kleckner et al., 2015; Ring et al., 2015; Ring & Brener, 2018; Windmann et al., 1999). However, predictive coding model perspectives of interoception (Ainley et al., 2016; Seth & Friston, 2016) suggest heart rate beliefs are likely to be generated as a result of both explicit and implicit knowledge about heart rates, which may explain the associations found between heart rate beliefs and heartbeat detection ability.

Finally, VR driving simulations have limited validity for more ambiguous real-world driving tasks in semi-automated vehicles. As with other VR dilemmas studies, participants were presented with a highly improbable level of certainty about the outcomes of taking harmful action. Although a systematic analysis of this issue has yet to be conducted; longer decision-making timeframes allows greater cognitive processing which can influence harm-acceptance decisions (Suter & Hertwig, 2011). Other VR driving dilemma studies have used shorter response timeframes than the current study (Ju et al., 2016; Skulmowski et al., 2014; Sütfeld et al., 2017) but were not investigating a collaborative decision-making situation with a vehicle AI. We aimed to provide a greater level of storytelling and immersion with the VR dilemmas used, and relatively fast-paced emerging collision events. Panic-driven behaviour is likely to be much more inconsistent and erratic and has been highlighted as a particular issue of moral dilemma studies like this (Pan & Slater, 2011). For example, in Dilemma 1 (incongruent), a majority of participants pressed the ‘override’ button before the collision warning was presented, despite receiving pre-training about the task. The early presses often coincided with an explosion from a vehicle behind and shows how fear-induced, knee-jerk behaviours can occur even in highly controlled hypothetical VR dilemmas. Some researchers are experimenting with a series of continuously emerging VR moral dilemmas that measure participants’ harmful actions in less predictable, more dynamic ‘apocalyptic’ scenarios (Cristofari & Guitton, 2014).

Conclusion

In the near future, humans will be required to supervise morally sensitive decisions of autonomous vehicle AI in dynamic collision situations which could have serious

consequences for passengers, pedestrians and other drivers. Although correspondence between moral decision-making in VR versus 'real-life' is likely an unattainable goal, measuring the emotional and physiological correlates of behaviour in VR moral dilemmas brings us closer to being able to draw conclusions about the transferability of lab-based moral judgment research to more realistic moral dilemma scenarios. Psychological (Eyal & Liberman, 2012) and physical distance (Cushman et al., 2012; Greene et al., 2001; Miller et al., 2014) are important in our emotional responses to harmful acts. It is possible that reducing the typicality of specific driving actions in these vehicles, such as replacing a foot-pedal with a button, may increase distance between driving actions and their outcomes. This may affect behaviours such as driving speed, braking distance, taking manual control of automation, or even relying more heavily on automation to make difficult decisions in real-life moral dilemma scenarios. We found anticipatory physiological responses (specifically electrodermal activity and pre-ejection period; PEP) influenced response time to carry out harmful utilitarian action in VR moral dilemma driving scenarios. The typicality of the harmful action (button-press versus foot-pedal) was associated with differing patterns of physiological reactivity and response times. Individual differences in interoceptive accuracy predicted different patterns of PEP reactivity and response time in the foot-pedal and button-press conditions. Heartbeat detection ability may contribute to different patterns of cardiovascular reactivity when harmful action is required to achieve a utilitarian outcome which may influence response time to act. However, we did not find that the other parameters of interoception, beyond interoceptive accuracy, were significant in the relationship between physiological arousal and moral judgments. People's tendency to notice bodily sensations, their awareness of their interoceptive accuracy and gastric sensitivity appeared unrelated to the moral behaviour outcomes in this preliminary study. Since Damasio's (1996) pivotal Somatic Marker Hypothesis, exploring individual differences in interoception has been notably absent in morality research and warrants much further investigation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

H.B. conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper. S.G.R.B. contributed to the design and building of the Virtual Reality environments.

Acknowledgments

We thank Dr Susanna Martin and Dr Mary Nikolaidou (University of Bath) for lab support and setting up of Virtual Reality computer. We thank Tiffany Yip, Natalie Tham and Emilia Tor (University of Bath) for support with data collection. We thank Dr Marc Holmes and Ted Thomas (University of Swansea) for technical support with the VR environments. We thank Dr Kathryn Francis (University of Bradford), Dr Harry Farmer (University of Greenwich), Dr Chris Bevan (University of Bristol) and Dr Dave Green for useful discussions.

References

Ainley, V., Apps, M. A. J., Fotopoulou, A., & Tsakiris, M. (2016). "Bodily precision": A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0003

Ainley, V., & Tsakiris, M. (2013). Body Conscious? Interoceptive Awareness, Measured by Heartbeat Perception, Is Negatively Correlated with Self-Objectification. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0055568

Aldao, A. (2013). The Future of Emotion Regulation Research: Capturing Context. *Perspectives on Psychological Science*, *8*(2), 155–172.

https://doi.org/10.1177/1745691612459518

Anderson, E. R., & Hope, D. A. (2009). The relationship among social phobia, objective and perceived physiological reactivity, and anxiety sensitivity in an adolescent population. *Journal of Anxiety Disorders*, *23*(1), 18–26. https://doi.org/10.1016/j.janxdis.2008.03.011

Anderson, S. W., Bechara, A., Damásio, H., Tranel, D., & Damásio, A. R. (2013). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Social Neuroscience: Key Readings*, *9780203496*(11), 29–40. https://doi.org/10.4324/9780203496190

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159–162. https://doi.org/10.1016/j.tics.2005.02.002

BIOPAC Systems, I. (2015). *Acknowledge Demo Guide*.

BIOPAC Systems, I. (2020). *MP160 BIOPAC Sytems*. https://www.biopac.com/product/mp150-system-221/

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *J.P. Forgas (Ed.) Feeling and thinking: The role of affect in social cognition* (Ed., pp. 59–82). Cambridge University Press.

Blascovich, J., & Mendes, W. B. (2010). Social Psychophysiology and Embodiment. *Handbook of Social Psychology*, 194–227. https://doi.org/10.1002/9780470561119.socpsy001006

Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2008). Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *17*(4), 376–391. https://doi.org/10.1162/pres.17.4.376

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Braithwaite, J. J., & Watson, D. G. (2015). Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs). *Technical Research Note.* Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham ©

Braithwaite, J. J., Watson, D. G., Robert, J., & Mickey, R. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. …. *Technical Report.* Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham © https://doi.org/10.1017.S0142716405050034

Brown, H., Proulx, M. J., & Fraser, D. S. (2020). Hunger Bias or Gut Instinct ? Responses to Judgments of Harm Depending on Visceral State Versus Intuitive Decision-Making. *Frontiers in Psychology: Perception Science*, *11*(September), 1–16. https://doi.org/10.3389/fpsyg.2020.02261

Brown, H., Fraser, D. S., & Proulx, M. J. (2020a). *Does modality of harmful action in a semi-autonomous vehicle influence responses to virtual-reality moral dilemmas?* Open Science Framework. https://osf.io/vhbdy.

Brown, H., Proulx, M. J., & Fraser, D. S. (2020b). *Do individual differences in interoception influence the relationship between physiological arousal and moral judgments of harm? [Unpublished thesis]*. University of Bath.

Byrd, N., & Conway, P. (2019). Not all who ponder count costs: Arithmetic reflection predicts utilitarian tendencies, but logical reflection predicts both deontological and utilitarian tendencies. *Cognition*, *192*(June), 103995. https://doi.org/10.1016/j.cognition.2019.06.007

Carmona-Perera, M., Reyes del Paso, G. A., Pérez-García, M., & Verdejo-García, A. (2013). Heart rate correlates of utilitarian moral decision-making in alcoholism. *Drug and Alcohol Dependence*, *133*(2), 413–419. https://doi.org/10.1016/j.drugalcdep.2013.06.023

Clark, D. M., Salkovskis, P. M., Breitholtz, E., Westling, B. E., Öst, L. G., Koehler, K. A., Jeavons, A., & Gelder, M. (1997). Misinterpretation of body sensations in panic disorder. *Journal of Consulting and Clinical Psychology*, *65*(2), 203–213. https://doi.org/10.1037/0022-006X.65.2.203

Craig, B. A. D. (2015). *How do you Feel?:An Interoceptive Moment with Your Neurobiological Self*. Princeton University Press.

Cristofari, C., & Guitton, M. J. (2014). Surviving at any cost: Guilt expression following extreme ethical conflicts in a virtual setting. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0101711

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*(2), 189–195. https://doi.org/10.1038/nn1176

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: the aversion to harmful action. *Emotion (Washington, D.C.)*, *12*(1), 2–7. https://doi.org/10.1037/a0025071

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81–94.

Damasio, A. R. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

De Berardis, D., Campanella, D., Gambi, F., La Rovere, R., Sepede, G., Core, L., Canfora, G., Santilli, E., Valchera, A., Mancini, E., Salerno, R. M., Moschetta, F. S., & Ferro, F. M. (2007). Alexithymia, fear of bodily sensations, and somatosensory amplification in young outpatients with panic disorder. *Psychosomatics*, *48*(3), 239–246. https://doi.org/10.1176/appi.psy.48.3.239

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*(26). https://doi.org/10.3389/fpsyg.2015.00026

Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, *30*(1), 1–11. https://doi.org/10.1016/j.cpr.2009.08.008

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., & Dalgleish, T. (2010). Listening to Your Heart: How Interoception Shapes Emotion Experience and Intuitive Decision Making. *Psychological Science*, *21*(12), 1835–1844. https://doi.org/10.1177/0956797610389191

Eichler, S., Katkin, Edward, S., Blascovich, J., & Kelsey, R. M. (1987). Cardiodynamic factors in heartbeat detection and the experience of emotion [Abstract]. *Psychophysiology*, *24*, 587.

Eichler, S., & Katkin, E. S. (1994). The relationship between cardiovascular reactivity and heartbeat detection. *Psychophysiology*, *31*(3), 229–234. https://doi.org/10.1111/j.1469-8986.1994.tb02211.x

Eyal, T., & Liberman, N. (2012). Morality and psychological distance: A construal level theory perspective. *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, 185–202. https://doi.org/10.1037/13091-010

Faraway, J. J. (2015). *Linear models with R* (2nd ed.). CRC Press.

Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., & Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in Psychology*, *6*(JUN), 1–26. https://doi.org/10.3389/fpsyg.2015.00763

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2018). Human Decisions in Moral Dilemmas are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-018-0020-x

Feldman Barrett, L. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsw154

Feldman Barrett, L., & Simmons, W. K. (2015). Interoceptive predictions in the brain. In *Nature Reviews Neuroscience* (Vol. 16, Issue 7). https://doi.org/10.1038/nrn3950

Feldman Barrett, L., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional. *J Pers Soc Psychol.*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A. K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human Computer Studies*, *82*, 48–56. https://doi.org/10.1016/j.ijhcs.2015.05.004

Foot, P. (2003). The Problem of Abortion and the Doctrine of the Double Effect. *Virtues and Vices*, *5*, 19–32. https://doi.org/10.1093/0199252866.003.0002

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating Moral Actions: An Investigation of Personal Force in Virtual Moral Dilemmas. *Scientific Reports*. https://doi.org/10.1038/s41598-017-13909-9

Francis, Kathryn B., Howard, C., Howard, I. S., Gummerum, M., Anderson, G. G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0164374

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: Interoceptive awareness facilitates reappraisal. *Social*

Cognitive and Affective Neuroscience, *8*(8), 911–917. https://doi.org/10.1093/scan/nss089

Garfinkel, S. N., & Critchley, H. D. (2016). Threat and the Body: How the Heart Supports Fear Processing. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.10.005

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 65–74. https://doi.org/10.1016/j.biopsycho.2014.11.004

Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology*, *114*, 117–126. https://doi.org/10.1016/j.biopsycho.2015.12.003

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, *45*(3), 581–584. https://doi.org/10.1016/j.jesp.2009.01.003

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371. https://doi.org/10.1016/j.cognition.2009.02.001

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105 LP – 2108. http://science.sciencemag.org/content/293/5537/2105.abstract

Gu, J., Zhong, C. B., & Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *Journal of Experimental Psychology: General*, *142*(2), 307–312. https://doi.org/10.1037/a0029549

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2013). Emotional regulation and bodily sensation: Interoceptive awareness is intact in borderline personality

disorder. *Journal of Personality Disorders*, *27*(4), 506–518. https://doi.org/10.1521/pedi_2012_26_049

Heeter, C., Day, T., & Cherchiglia, L. (2020). Interoceptive Awareness: The 'Being' Dimension of 'Being There' in Virtual Worlds. *Interacting with Computers*, *32*(1), 1–16. https://doi.org/10.1093/iwc/iwaa001

Herbert, B. M., Ulbrich, P., & Schandry, R. (2007). Interoceptive sensitivity and physical effort: Implications for the self-control of physical load in everyday life. *Psychophysiology*, *44*(2), 194–202. https://doi.org/10.1111/j.1469-8986.2007.00493.x

Ju, U., Kang, J., & Wallraven, C. (2016). Personality differences predict decision-making in an accident situation in virtual driving. *Proceedings - IEEE Virtual Reality*, *2016-July*, 77–82. https://doi.org/10.1109/VR.2016.7504690

Kant, I. (2018). Groundwork of the Metaphysics of Morals (1785). *Business Ethics: The Big Picture*, 18.

Katkin, E. S. (1985). Blood, sweat, and tears: Individual differences in autonomic self-perception. *Psychophysiology*, *22*(2), 125–137.

Kever, A., Pollatos, O., Vermeulen, N., & Grynberg, D. (2015). Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. *Personality and Individual Differences*, *87*. https://doi.org/10.1016/j.paid.2015.07.014

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, *69*(March 2017), 66–73. https://doi.org/10.1016/j.apergo.2017.12.016

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology*, *52*(11), 1432–1440. https://doi.org/10.1111/psyp.12503

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.4324/9781315853161

Kruschwitz, J. D., Kausch, A., Brovkin, A., Keshmirian, A., Paulus, M. P., Goschke, T., & Walter, H. (2019). Self-control is linked to interoceptive inference: Craving regulation and the prediction of aversive interoceptive states induced with inspiratory breathing load. *Cognition*, *193*(July), 104028. https://doi.org/10.1016/j.cognition.2019.104028

Lacey, B. C., & Lacey, J. I. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. *American Psychologist*, *33*(2), 99–113. https://doi.org/10.1037/0003-066X.33.2.99

McDonald, M. M., Defever, A. M., & Navarrete, C. D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior*, *38*(6), 770–778. https://doi.org/10.1016/j.evolhumbehav.2017.06.001

Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PLoS ONE*, *13*(12), 1–12. https://doi.org/10.1371/journal.pone.0208034

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the Unexpected: Physiological Responses During Social Interactions With Expectancy-Violating Partners. *Journal of Personality and Social Psychology*, *92*(4), 698–716. https://doi.org/10.1037/0022-3514.92.4.698

Mill, J. S. (1998). *Utilitarianism* (R. Crisp,). Oxford University Press.

Miller, L.C., Murphy, R. and Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology*, *41*(2), 397.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion (Washington, D.C.)*, *14*(3), 573—587. https://doi.org/10.1037/a0035361

Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–1899. https://doi.org/10.1162/jocn.2009.21367

Nakamura, J., & Csikszentmihalyi, M. (2005). The Concept of Flow. In *Handbook of Positive Psychology* (pp. 89–105).

Naqvi, N. H., & Bechara, A. (2006). Skin conductance: A psychophysiological approach to the study of decision making. In C. Senior, T. Russell & M. Gassaniga (Ed.), *Methods in mind* (pp. 103–122). https://doi.org/10.1088/2058-7058/26/05/35

Newlin, D. B., & Levenson, R. W. (1979). Pre-ejection Period: Measuring Beta-adrenergic Influences Upon the Heart. *Psychophysiology*, *16*(6), 546–552. https://doi.org/10.1111/j.1469-8986.1979.tb01519.x

OMRON, I. (n.d.). *OMRON M3 Comfort*. www.omron-healthcare.co.uk/blood-pressure-monitors/m3_comfort_2.html

Pan, X., & Slater, M. (2011). Confronting a Moral Dilemma in Virtual Reality: A Pilot Study. *Proceedings of the 25th BCS Conference on Human-Computer Interaction SE - BCS-HCI '11*, 46–51. citeulike-article-id:13326879%5Cnhttp://portal.acm.org/citation.cfm?id=2305326

Parton, D. M., & McGinley, J. J. (2019). At the heart of harm: Cardiac substrates of action-based aversion to harm. *Personality and Individual Differences*, *151*(July 2018), 109497. https://doi.org/10.1016/j.paid.2019.07.007

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2020). Reasoning Supports Utilitarian Resolutions to Moral Dilemmas Across Diverse Measures. *Journal of Personality and Social Psychology*, 1–42. https://doi.org/10.1037/pspp0000281

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure & Function*, *214*(5–6), 451–463. https://doi.org/10.1007/s00429-010-0258-9

Plotnick, R. (2018). At the Interface : The Case of the Electric Push Button , 1880 – 1923. *Technology and Culture*, *53*(4), 815–845.

Pollatos, O., Herbert, B. M., Kaufmann, C., Auer, D. P., & Schandry, R. (2007a).

Interoceptive awareness, anxiety and cardiovascular reactivity to isometric exercise. *International Journal of Psychophysiology*, *65*(2), 167–173. https://doi.org/10.1016/j.ijpsycho.2007.03.005

Pollatos, O., Herbert, B. M., Matthias, E., & Schandry, R. (2007b). Heart rate response after emotional picture presentation is modulated by interoceptive awareness. *International Journal of Psychophysiology*, *63*(1), 117–124. https://doi.org/10.1016/J.IJPSYCHO.2006.09.003

Pollatos, O., Matthias, E., & Keller, J. (2015). When interoception helps to overcome negative feelings caused by social exclusion. *Frontiers in Psychology*, *6*(JUN), 1–8. https://doi.org/10.3389/fpsyg.2015.00786

Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion*, *18*(7), 1009–1023. https://doi.org/10.1037/emo0000413

Ring, C., & Brener, J. (2018). Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. *Psychophysiology*, *February*, 1–10. https://doi.org/10.1111/psyp.13084

Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: A cautionary tale about interoceptive awareness. *Biological Psychology*, *104*, 193–198. https://doi.org/10.1016/J.BIOPSYCHO.2014.12.010

Schandry, R. (1981). Heart Beat Perception and Emotional Experience. *Psychophysiology*, *18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Scheepers, D., de Wit, F., Ellemers, N., & Sassenberg, K. (2012). Social power makes the heart work more efficiently: Evidence from cardiovascular markers of challenge and threat. *Journal of Experimental Social Psychology*, *48*(1), 371–374. https://doi.org/10.1016/j.jesp.2011.06.014

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality & Social Psychology Bulletin*, *34*(8), 1096–1109. https://doi.org/10.1177/0146167208317771

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*,

371(1708). https://doi.org/10.1098/rstb.2016.0007

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *3*(JAN), 1–16. https://doi.org/10.3389/fpsyg.2011.00395

Sherwood, A., Dolan, C. A., & Light, K. C. (1990). Hemodynamics of Blood Pressure Responses During Active and Passive Coping. *Psychophysiology*, *27*(6), 656–668. https://doi.org/10.1111/j.1469-8986.1990.tb03189.x

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*(December), 1–16. https://doi.org/10.3389/fnbeh.2014.00426

Slater, M., Steed, A., & Usoh, M. (1995). *The Virtual Treadmill: A Naturalistic Metaphor for Navigation in Immersive Virtual Environments*. *January*, 135–148. https://doi.org/10.1007/978-3-7091-9433-1_12

Slater, M., Usoh, M., & Steed, A. (2000). Taking steps: The Influence of a Walking Technique on Presence in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, *2*(3), 201–219. https://doi.org/10.1049/me:20000204

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE):Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=9716&login.asp&site=ehost-live&scope=site

Solovey, E. T., Zec, M., Perez, E. A. G., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: Two field studies. *Conference on Human Factors in Computing Systems - Proceedings*, 4057–4066. https://doi.org/10.1145/2556288.2557068

Sood, S., & Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organizational Behavior and Human Decision Processes*, *98*(2), 144–154. https://doi.org/10.1016/j.obhdp.2005.05.005

Spielberger, C. D., & Gorsuch, R. L. (1983). *The state trait anxiety inventory for adults: sampler set manual, test, scoring key*. Redwood City Calif: Mind Garden.

Starcke, K., Ludwig, A. C., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral judgment. *Judgment and Decision Making*, *7*(1), 61–68. https://doi.org/10.17185/duepublico/45052

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458. https://doi.org/10.1016/j.cognition.2011.01.018

Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: Applicability of value-of-life-based models and influences of time pressure. *Frontiers in Behavioral Neuroscience*, *11*(July), 1–13. https://doi.org/10.3389/fnbeh.2017.00122

Takahashi, T., Ikeda, K., Ishikawa, M., Kitamura, N., Tsukasaki, T., Nakama, D., & Kameda, T. (2005). Anxiety, reactivity, and social stress-induced cortisol elevation in humans. *Neuroendocrinology Letters*, *26*(4), 351–354.

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, Physiological, and Behavioral Effects of Threat and Challenge Appraisal. *Journal of Personality and Social Psychology*, *65*(2), 248–260. https://doi.org/10.1037/0022-3514.65.2.248

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*. https://doi.org/10.1037/a0018963

Turner, M. J., Jones, M. V., Sheffield, D., Slater, M. J., Barker, J. B., & Bell, J. J. (2013). Who thrives under pressure? Predicting the performance of elite academy cricketers using the cardiovascular indicators of challenge and threat states. *Journal of Sport and Exercise Psychology*, *35*(4), 387–397. https://doi.org/10.1123/jsep.35.4.387

Uijong, J., Kang, J., & Wallraven, C. (2019). You or Me? Personality Traits Predict Sacrificial Decisions in an Accident Situation. *IEEE Transactions on Visualization and Computer Graphics*, *25*(5), 1898–1907. https://doi.org/10.1109/TVCG.2019.2899227

Ullén, F., de Manzano, Ö., Theorell, T., & Harmat, L. (2013). The Physiology of Effortless

Attention: Correlates of State Flow and Flow Proneness. *Effortless Attention*, *April*, 205–218. https://doi.org/10.7551/mitpress/9780262013840.003.0011

Van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P. C., Schulz, A., & Herbert, B. M. (2016). The water load test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PLoS ONE*, *11*(9), 1–14. https://doi.org/10.1371/journal.pone.0163574

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, *24*(5), 2274–2291. https://doi.org/10.1016/j.chb.2007.11.002

Weiss, S., Sack, M., Henningsen, P., & Pollatos, O. (2014). On the interaction of self-regulation, interoception and pain perception. *Psychopathology*, *47*(6), 377–382. https://doi.org/10.1159/000365107

Wiens, S. (2005). Interoception in emotional experience. *Current Opinion in Neurology*, *18*(4), 442–447. https://doi.org/10.1097/01.wco.0000168079.92106.99

Windmann, S., Schoneck, O. W., Frohlig, G., & Maldener, G. (1999). Dissociating beliefs about heart rates and actual heart rates in patients with cardiac pacemakers. *Psychophysiology*, *36*(3), 339–342.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to Ventromedial Prefrontal Cortex Impairs Judgment of Harmful Intent. *Neuron*, *65*(6), 845–851. https://doi.org/10.1016/j.neuron.2010.03.003

Zeng, J., Zuo, X.-L., Li, Y.-Q., Wei, W., & Lv, G.-P. (2007). Tegaserod for dyspepsia and reflux symptoms in patients with chronic constipation: an exploratory open-label study. *European Journal of Clinical Pharmacology*, *63*(6), 529–536. https://doi.org/10.1007/s00228-007-0287-3

Study 3 Postface

Paper:

Brown, H., Fraser, D. S., Brown, S. G. R., & Proulx, M. J. (2020). *Moral dilemmas in a semi-autonomous vehicle: the effects of physiological arousal, interoception and typical versus non-typical driving actions in VR driving dilemmas [Unpublished thesis]*. University of Bath.

Although this study sample was smaller than planned due to the coronavirus pandemic, the findings introduce some interesting avenues for future research exploring the link between arousal, moral behaviour and interoception. Using four different measures of physiological arousal provided insights into the utility of each of these parameters for predicting moral behaviour in VR moral dilemmas. As Parton and McGinley (2019) found, pre-ejection period (PEP) and not our index of systemic vascular resistance (SVR; Cushman et al., 2012) was associated with individual differences in interoceptive accuracy and group differences (button press versus foot-pedal) in response time. These findings are in agreement with evidence showing PEP is a more appropriate measure of sympathetic cardiovascular reactivity in active-coping tasks (Sherwood et al., 1990) such as VR moral dilemmas exploring action-based harm aversion (Parton & McGinley, 2019).

Electrodermal activity showed positive correlation with response time in both the foot-pedal and button-press conditions. In the Iowa Gambling Task, healthy controls [compared to patients with ventromedial damage] tend to show an anticipatory electrodermal response prior to taking a card from a 'bad' deck, which does not happen for 'good' decks (Bechara et al., 1996). This anticipatory response also happens for controls before endorsing 'personal' harmful acts in moral dilemmas, but not for patients with ventromedial lesions (Moretto et al., 2010). Damasio et al (1996) suggested skin conductance responses, may serve as an anticipatory 'alert' response [initiated from the ventromedial cortex] to the prospect of a negative outcome, which can bias decision-making. Therefore, a stronger, more sustained electrodermal response may indicate an inhibitory 'alert' response that prevented people from acting quickly. The fact that electrodermal activity showed the same relationship with response time in the foot-pedal and button-press conditions, suggests this may have been an indiscriminate 'alert', ahead of anticipating a bad outcome that was not influenced by the type of harmful action used. Electrodermal activity may be associated with a more

generalised warning associated with negative outcomes, as opposed to the actions, in moral dilemmas. For example, Patil et al (2014) found that skin conductance responses were higher in VR dilemmas that were 'impersonal' (akin to the Trolley problem), compared to text-based versions. It is possible that the particularly arousing nature of VR moral dilemmas may increase hyper-vigilance towards taking actions that are morally wrong, as the consequences and responsibility for the outcomes of harm are more socially visible (Patil et al., 2014).

The finding that interoceptive accuracy predicted a lengthening of PEP reactivity (reduced force of myocardial contractility) prior to harmful action on incongruent trials relative to congruent trials was surprising. Although it is possible that people experienced the congruent trials as more 'stressful', we expected most people to demonstrate heightened sympathetic reactivity on incongruent trials - as the condition necessitating harmful action to save 'lives' of passengers. Earlier studies have found that better heartbeat detection is associated with faster/shorter PEP reactivity to aversive stimuli (Eichler et al., 1987; Eichler & Katkin, 1994) and physical stress tasks (Pollatos et al., 2007a). Whereas, our results suggest interoceptive accuracy predicted longer PEP reactivity, which is more typically associated with motivational 'challenge' states on active coping tasks (Tomaka et al., 1993). This pattern of results could be due to fundamental individual differences in people's cardiovascular reactivity to active coping tasks (Sherwood et al., 1990). Alternatively, it could be associated with a reduced 'effort' among these people to carry out the harmful actions. Herbert et al (2007) found that people higher on interoceptive accuracy exerted less physical effort during exercise, proposedly due to their superior ability to perceive bodily signals. Physiological measures such as PEP may not show the expected patterns of results for better heartbeat perceivers in tasks requiring physical effort, which may mean that the associations between physiological arousal and harmful moral action are not so straightforward. However, the relationship between IAc and response time was moderated by action-type and suggests that if better heartbeat detectors had a superior ability to regulate physical and cardiovascular effort when carrying out the actions, this did not apply equally to typical and non-typical actions. Relatedly, another study found that 'Consistency' (a measure of people's ability to focus on and stick to a small group of valuable goals) was associated with slower PEP on a mental effort task (Silvia et al., 2013). They speculatively suggested that people high on consistency may not appraise short term tasks as useful to achieve valuable goals, and therefore may deem them less important. Furthermore, reviews of cardiovascular parameters of challenge and threat suggest that systemic-vascular resistance

and cardiac output are stronger indicators of challenge and threat associated with the biopsychosocial model (Blascovich & Mendes, 2000), and measures of heart rate and PEP are more reliably associated with task engagement (Behnke & Kaczmarek, 2018; Seery, 2011). Seery (2011), suggested that previous work showing corresponding changes in HR and PEP alongside changes in SVR and CO, may be due to greater task engagement, which may have been conflated with 'challenge' states. Further work is needed to understand how PEP and physical exertion during moral behaviour experiments are associated with interoceptive accuracy and self-reported task engagement.

We did not find a significant association between trait or state anxiety and IAc in Study 3, contrary to prior research finding positive correlations between trait anxiety, heartbeat detection ability and cardiovascular reactivity (e.g. Critchley et al., 2004; Pollatos et al., 2007a). This may be due to the small sample size. Anxiety was also not associated (p>.05) with any of the physiological indexes in the pre-decision timeframe, suggesting that anxiety was not responsible for changes in cardiovascular reactivity prior to taking harmful action on incongruent trials. In Study 2, we found a negative association between trait anxiety and IAc, but trait anxiety did not influence any of the moderation effects of IAc in the relationship between physiological arousal and moral judgment. Consistent with this finding, Pollatos and colleagues (2007a) suggested that despite common variation among measures of heartbeat detection ability and anxiety, cardiovascular reactivity may be independently associated with these two variables.

The absence of a positive relationship between IAc and anxiety in Study 2 and 3, may be due to eligibility criteria stating that participants should not be experiencing any current mental health conditions. This may have reduced the opportunity for particularly high levels of anxiety within the sample. This criterion was consciously set to reduce the possibility of anxiety confounding our interoception analyses, and also due to the potentially distressing nature of the studies. The negative relationship between trait anxiety and IAc in Study 2, is consistent with evidence showing heartbeat detection ability is negatively associated with a range of measures of health anxiety, indicating that at least health-related forms of anxiety are associated with less accurate interoceptive sensitivity (Krautwurst et al., 2014). However, in Study 3, interoceptive sensibility was negatively associated with state anxiety, (p<.05), which challenges the notion that a greater tendency to notice bodily sensations is associated with heightened self-reported arousal. This may be due to the measure of interoceptive

sensibility used (Miller, Murphy & Buss, 1981). In Study 1 this measure of interoceptive sensibility positively correlated with self-reported mindfulness, which suggests this measure may sometimes capture more adaptive forms of bodily attention.

We did not find any associations between any of the other interoception measures and the moral behaviour measures. Action-based harm aversion has historically been associated with cardiovascular indexes of sympathetic arousal (Cushman et al., 2012; Parton & McGinley, 2019). This could explain why we only found associations between interoceptive accuracy, PEP and response time in Study 3 as this dimension of interoception is most likely to moderate the relationship between cardiovascular arousal and behaviour. Gastric interoception in particular may not play as pivotal role in moral behaviour experiments, as participants are not required to subjectively assess the moral acceptability of harmful action. Evaluating the acceptability of harmful actions retrospectively from an allocentric perspective (as in Study 2), could have provided opportunity to refer to embodied sensations of disgust previously found to be associated with judgments of ethical violations including harmful acts of others (Schnall et al., 2008; Vicario et al., 2018). As discussed in Study 1, it may be that egocentric harm-based moral judgments are less influenced by gastric sensations associated with disgust, that have been associated observations of 'bad' moral character (Giner-Sorolla & Chapman, 2017) and moral violations of others (Tracy et al., 2019). This could be due to the fundamental function of disgust as a defensive 'rejection' mechanism to protect the organism from a range of contaminating and violating external threats to the body and soul (Rozin, 2015). Disgust sensations may therefore be less useful to egocentric judgments of harm, as disgust towards immoral action would need to be self-directed. Self-directed emotions of shame and guilt are intuitively more influential for egocentric judgments of harm where you consider the self-relevant consequences of immoral action. The differentiation between the heart and stomach is interesting as it suggests sensations within the cardiac and gastrointestinal visceral systems may be discretely associated with different types of moral judgment.

Study 4 Preface

Paper:

Brown, H., Fraser, D. S., Farmer, H., & Proulx, M. J. (2020). *The salience of harm: the framing and harm-content of coronavirus media articles indirectly influences preventative and antisocial behaviours during the pandemic [Unpublished thesis].* University of Bath.

Draft paper submitted to:

Special issue: *Risk Perception, Communication, and Decision Making in the Time of COVID-19; Journal of Experimental Psychology: Applied*

Study 4 was conceived at the onset of the coronavirus pandemic. These exceptional circumstances provided a unique context to explore morally relevant behaviours during national lockdowns during April-May 2020. In the early days of the crisis in Europe there was an overwhelming amount of media content depicting traumatic stories and images of people who were suffering with and dying from coronavirus. These were often shocking and highly emotional stories, that evoked a strong imagery regarding the unknowable and indiscriminate threat of this new deadly disease. Some articles were clearly more 'harm-salient' in their descriptions of the pandemic, highlighting the threat of death and illness as a central concern that was likely to have a strong emotional impact on readers. People were also being told to make significant changes to their behaviour on a massive scale. Public health messages were emphasising the need to make personal sacrifices such as social distancing, in order to control the spread of the virus, ultimately to save lives and reduce the strain on health services. However, an early research study during coronavirus, found emotionally arousing media frames and content during this time was increasing emotional distress of readers (e.g. Tabri et al., 2020). Anxiety in particular has the potential to encourage preoccupation with selfish needs during the pandemic, as our sense of threat-perception is heightened (Barlow, 1988). We hypothesised that self-oriented distress may make people less inclined to consider the wider social implications of their behaviour, potentially reducing the likelihood that they would behave in a way that benefits a 'greater good' at a societal level. Although, interestingly, anxiety has also shown to be useful in facilitating some preventative behaviours during previous pandemics (e.g. Bults et al., 2011).

In addition, as this was a health-centric crisis, we were particularly interested in whether people's tendency to focus on or worry about interoceptive or bodily sensations influenced their interpretation and emotional and behavioural responses to news media about coronavirus.

We tested the effects of frame-type and harm-salience of coronavirus media on people's behavioural intentions and moral judgments during the pandemic and investigated whether individual differences in interoceptive sensibility and anxiety influenced these effects. We manipulated the amount of harm-content (high versus low) and also the frame-type (individual focus i.e. episodic, versus wider societal focus i.e. thematic) of fictitious coronavirus media articles. This time, we used the 'noticing', 'not worrying' and 'not distracting' subscales from the Multidimensional Interoceptive Awareness scale (Mehling et al., 2018) as opposed to the Body Consciousness scale (Miller, Murphy, & Buss, 1981) used in study 1-3. These subscales were more comprehensive and appropriate to assess tendencies to worry or catastrophise about bodily sensations, that would be most relevant to their responses to threatening health-related media. We limited the measure to these three subscales to reduce experimental load and to focus our hypotheses.

We developed novel moral dilemmas that were specifically related to the coronavirus pandemic, to explore whether moral judgments directly related to real-world moral dilemmas showed similar patterns to other generic moral dilemmas. Moral dilemmas have not previously been curated to reflect immediate real-world issues that participants may be personally affected by, which could increase emotional engagement with these tasks. The dilemmas were either personal (in the role of self) or impersonal (in the role of a healthcare worker) to ascertain whether adopting these roles facilitated different responses, and whether these responses predicted real-world behavioural intentions during the pandemic.

The role of media effects on moral judgment and behaviour has not been explored before in a pandemic context. Furthermore, the role of individual differences in interoception has not previously been investigated as a factor that could influence people's behavioural responses to emotionally arousing media content.

Statement of Authorship

This declaration concerns the article entitled:			
The salience of harm: the framing and harm-content of coronavirus media articles indirectly influences preventative and antisocial behaviours during the pandemic			
Publication status (tick one)			
Draft manuscript		**Submitted** x	**In review**		**Accepted**		**Published**	
Publication details (reference)	Draft paper submitted to: *Journal of Experimental Psychology: Applied*, Special Issue: *Risk Perception, Communication, and Decision Making in the Time of COVID-19* Authors: Helen Brown, Danaë Stanton Fraser, Harry Farmer, and Michael J. Proulx.		
Copyright status (tick the appropriate statement)			
I hold the copyright for this material	X	Copyright is retained by the publisher, but I have been given permission to replicate the material here	
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed the formulation of ideas, design of methodology, experimental work and presentation of data in journal format. The candidate (H.B.) conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper. H.F. provided support with data analysis and feedback on written paper.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	This information has been redacted for privacy reasons.	**Date**	05/11/2020

The salience of harm: the framing and harm-content of coronavirus media articles indirectly influences preventative and antisocial behaviours during the pandemic

Helen Brown [1,2], Danaë Stanton Fraser [2], Harry Farmer [3,4] and Michael Proulx [1]

[1] Crossmodal Cognition Lab, Department of Psychology, University of Bath

[2] CREATE Lab, Department of Psychology, University of Bath

[3] School of Human Sciences, University of Greenwich

[4] Institute of Lifecourse Development, University of Greenwich

Author Note

Research funded by University of Bath. This studentship runs alongside UK EPSRC grant Virtual Realities: Immersive Documentary Encounters (ref: EP/P025595/1).

Correspondence concerning this article should be addressed to Helen Brown, hrhb20@bath.ac.uk.

Abstract

Media coverage of the COVID-19 pandemic is vital for providing information to the public to encourage informed and united action that can reduce the spread of the virus. Media frames can affect people's emotions and exert influence by accentuating certain issues over others, increasing the accessibility or saliency of these issues. Episodic frames are more individualistic, surround specific events and are more likely to use evocative engagement strategies. Thematic frames lie in the public realm, integrate events and stories into one fundamental issue and provide audiences with useful background knowledge. This online experiment (n=151) investigated whether frame-type (episodic versus thematic) and the harm-salient content of coronavirus media stories influenced feelings of anxiety, health-relevant behavioural intentions and moral judgments. The harm-content and the frame-type indirectly influenced people's intention to carry out preventative and anti-social behaviours by shaping subjective appraisals of harm-salience. Anxiety moderated whether people's appraisals of harm-salience translated into preventative or antisocial behaviours. Moral judgments were less affected by the news coverage and in the main did not correlate with behavioural intentions, suggesting moral judgments regarding the treatment of coronavirus patients are not informative for predicting real-world behaviours during the pandemic.

Keywords: COVID-19, media frames, moral behaviour, anxiety, interoception, harm, salience.

Introduction

The importance of media coverage during a pandemic

COVID-19 is a novel form of the coronavirus disease primarily affecting the lungs and airways which continues to be a threat to public health worldwide while we await a vaccine. Although less deadly than other epidemic diseases such as SARS and Ebola, coronavirus is extremely contagious. From its origins in Wuhan, China, in 2019, coronavirus had spread globally within two months and at the time of writing this article has cost the lives of over 1.4 million people worldwide. Media coverage during previous pandemics has shown to be vital for providing information to the public to encourage informed and united action that can reduce viral spread (Bults et al., 2011; Yan, Tang, Gabriele, & Wu, 2016). Equally, coverage regarding 'threats' is likely to be regarded as more newsworthy during a pandemic (Chang, 2012; Klemm, Das, & Hartmann, 2016) and the media has been criticised for dramatising or sensationalising the deadliness and threat of previous pandemics (Klemm et al., 2016) which has the potential to stir up public anxieties (Bodas, Siman-Tov, Peleg, & Solomon, 2015; Chang, 2012; Tabri, Hollingshead, & Wohl, 2020).

Media framing effects

Media frames are a category of rhetorical devices used by media outlets that can influence people's perceptions of risk during health crises (Chang, 2012), by shaping how people interpret (e.g. Hart, 2011; Iyengar & Simon, 1993) and respond to public issues (Price, Tewksbury, & Powers, 1997; Maddux & Rogers, 1983). Gamson and Modigliani describe media framing as "a central organizing idea or story line that provides meaning to an unfolding strip of events, weaving a connection among them." (1987, p.143). This study is situated in the literature of emphasis framing effects, whereby the media influences opinions by accentuating certain issues or themes over others, thus increasing the accessibility of these issues in the minds of audiences (Druckman, 2001). Agenda-setting theory is also relevant and proposes a strong relationship between the emphasis the media places on certain issues and how important audiences evaluate these issues to be (see Scheufele & Iyengar, 2012 for a review). In particular, we were interested in the effects of episodic and thematic frames of coronavirus news coverage. Episodic frames are more individualistic in nature, surround specific cases or events and are more likely to use evocative engagement strategies such as shocking images or headlines. In contrast, thematic frames lie in the public realm, integrate

events and stories into one fundamental issue and seek to provide audiences with background knowledge (Iyengar, 1994; Nitz & West, 2004).

The use of episodic versus thematic frames can shape perceptions of who is responsible for dealing with societal issues and influence how likely we are to take personal action to resolve social problems. For example, when issues of poverty are framed in a thematic format (general statistics and outcomes) people attribute more responsibility for poverty to the government and wider society, whereas episodic frames (a case of one impoverished individual) encourage people to assign responsibility to that individual (Iyengar, 1994). Frames can be persuasive because people's attitudes or judgments are often influenced by how easily we can access certain types of information (Tversky & Kahneman, 1974). For example, narratives describing an experience of one individual victim (episodic) are more effective at encouraging charitable donations than a story involving numerous victims (Kogut & Ritov, 2005a, 2005b). Helping the one person identified appeals to the helper as we typically get greater satisfaction from helping a greater proportion of identified victims, in this case, helping 1 identified victim equates to helping 100% of identified victims (Kogut & Ritov, 2005a, 2005b).

Media frames are also purposefully used to emotionally affect audiences as a means to seduce, shock or influence them (Klemm et al., 2016). Our emotional reactions can then produce different behavioural responses as we try to gain congruency between what we feel and what we think (Nerb & Spada, 2001). Episodic frames are characteristically more emotionally arousing than thematic frames, as people typically have stronger emotional responses towards identified people compared to unidentified people in a story (Small & Loewenstein, 2003). Hart (2011) suggests that individuals represent discrete and coherent psychological entities which provide a more understandable narrative for interpretation. Consistent with this, Aarøe (2011) and Gross (2008) found episodic frames were more effective at inducing emotional responses towards the story's subject, than thematic frames. However, thematic frames were found to be more persuasive when no emotional response was present (Aarøe, 2011) or when accounting for emotional effects of the frames (Gross, 2008). People reported feeling more distress when reading stories about one identified victim, compared to multiple unidentified victims (Kogut & Ritov, 2005a), and gave more monetary contributions (Kogut & Ritov, 2005a, 2005b). In contrast, Hart (2011) found feelings of anxiety and worry triggered by media coverage stimulated behaviour change and policy

support, but a concern for victims did not. Hart proposed that frames which evoked feelings of self-focused anxiety or worry inspired self-serving 'prosocial' behaviour as an attempt to reduce negative feelings e.g. by donating to an associated cause.

Harm salience of news coverage

Although episodic frames are typically more affecting than thematic frames, some thematically framed narratives are inherently more passionate or shocking through their use of dramatic statistics or emotional language, which can make messages more memorable (Bolls, Lang, & Potter, 2001; Uribe & Gunter, 2007). Therefore, in addition to the effects of episodic and thematic frames, we sought to explore the effects of harm-salience, as an authored characteristic of news stories that could be particularly emotionally arousing during this pandemic. We conceptualised harm salience as the prominence of harm and suffering of people depicted within an article, which could be a property of both episodic and thematic stories about people who have been profoundly affected by the pandemic. In virtual reality research, Patil and colleagues (2014) found that graphically representing the number of people who could be harmed in hypothetical moral dilemmas increased people's physiological arousal and the likelihood they would carry out an action in the interests of saving the most victims. The authors proposed, the contextual saliency of the potential victims led people to place more value on the potential harmful outcomes of their decisions. The saliency of harm to others can also be exaggerated in written media, as both issue saliency and evaluative tone of media coverage are key factors dictating the direction and emotional strength of a media message (Sheafer, 2007). We suggest that articles using descriptive language that evokes stronger visual representations of harm or suffering of others affected by the pandemic, would be more emotionally arousing and accentuate the harmful consequences of the virus as a central issue, compared to less harm-salient articles.

Media coverage, anxiety and behaviour during a pandemic

As well as emphasising the harmful consequences of coronavirus, harm-salient narratives may also generate emotional distress if readers consider the risks and consequences of the coronavirus in their own lives. The social amplification of risk model (Kasperson et al., 2016) proposes communication of risk via news media, can selectively report some information in a way to increase or reduce people's perceptions of risk, irrespective of the accuracy of the coverage. Media coverage is classified as dramatizing, sensationalist, or 'hype' if it disproportionately draws attention to a threat, exaggerates risk (Vasterman,

Yzermans, & Dirkzwager, 2005) or predominantly discusses a threat using emotional or arousing content, as opposed to factual content (see Klemm et al., 2016 for review). Consistent with this, media frames that emphasise coronavirus as an existential threat have shown to increase anxiety in readers (Tabri et al., 2020). Anxiety has shown an inverse relationship with empathy, which is likely due to the fact that while empathy requires a consideration of other's needs, anxiety facilitates a preoccupation with selfish needs (Deardorff et al., 1977). Moreover, threatening or shocking stories of others have been shown to simultaneously increase anxiety and reduce empathy-related distress (Negd, Mallan, & Lipp, 2011). Anxiety increases threat-perception, makes us think in more simplistic ways (Barlow, 1988) and has been associated with unethical behaviours (e.g. Kouchaki & Desai, 2015). Reminding people of their own mortality can also increase people's self-oriented defences to avoid feelings of anxiety (Greenberg et al., 2003) which can impair moral decision-making (Trémolière, Neys, & Bonnefon, 2012) and influence health-related behaviours (Goldenberg & Arndt, 2008; Grover & Miller, 2014).

Conversely, epidemiological evidence suggests that anxiety may actually be useful for public health during a pandemic, as higher levels of anxiety were associated with more preventative measures to keep well during the Influenza pandemic in the Netherlands (Bults et al., 2011), during the swine-flu epidemic (Jones & Salathé, 2009) and during outbreaks of SARS in Hong Kong (Lau, Yang, Tsui, & Kim, 2003; Leung et al., 2005). Together, these findings suggest that media frames generating anxiety in readers may encourage egoistically driven preventative behaviours to keep oneself safe from coronavirus but may concurrently reduce motivations to carry out behaviours that serve the interests of other people. We explored whether anxiety after reading a coronavirus news article increased preventative behaviours but weakened people's capacity to carry out prosocial behaviours and/or avoid carrying out antisocial behaviours.

In addition, some people are more prone to noticing, paying attention to and/or worry about internal bodily sensations which can predispose them to catastrophising, anxiety and health-related concerns (Clark et al., 1997; Domschke et al., 2018). These dispositional tendencies represent subconstructs of interoceptive sensibility (Mehling et al., 2018) which could influence how people react to a global health crisis but to our knowledge has not been explored in this context before. Interoception refers to our perception of sensations originating in the body that accompany regulatory processes associated with homeostasis, such as cardiac or gastrointestinal feelings originating in the body (Craig, 2015). Some

people are more sensitive to changes in bodily sensations than others (Garfinkel & Critchley, 2013). Therefore, understanding individual differences in how people typically respond to somatic sensations potentially resembling symptoms of illness, discomfort or pain during this time could provide insight into how media communications about the harmful outcomes of coronavirus, may influence emotional and behavioural reactions to this media.

Do moral judgments predict behavioural responses to news coverage?

Finally, whether or not people choose to carry out behaviours that could prevent the spread of the virus or help others during the pandemic has moral implications. Collectively, we make sacrifices to our social lives, our time spent outdoors, and how often we visit friends and family, to ultimately contain the virus and prevent a rising death toll. As feelings of anxiety can make us more prone to focus on selfish needs (Deardorff et al., 1977) which can influence moral decision-making (Sarlo, Lotto, Rumiati, & Palomba, 2014; Zhao, Harris, & Vigo, 2016), we explored whether people's moral judgments regarding coronavirus moral dilemmas predicted their intentions to conduct themselves in ways that benefit public health in general (e.g. social distancing), specifically benefit others (e.g. volunteering) or specifically benefit themselves (e.g. bulk buying). There is growing research to suggest that people's hypothetical moral judgments often diverge from their moral behaviours when faced with the same scenarios in more realistic environments (e.g. Francis et al., 2016), but this judgment-behaviour discrepancy has not been explored in a pandemic context.

Present study

A primary aim of this study was to investigate the influence of frame-type and harm-salience of coronavirus news coverage on readers' anxiety, moral judgments and behavioural intentions associated with the COVID-19 pandemic.

Hypotheses

H1. Episodic frames and high harm-content articles will increase anxiety levels.

H2. People higher in interoceptive sensibility will demonstrate greater anxiety after reading the articles, independent of the effect of frame-type and harm-content.

H3. Episodic frames and a greater amount of harm-content in the news articles will induce stronger subjective representations of harm and suffering of their subjects.

H4. Frame-type and harm content will influence behavioural intentions by enhancing appraisals of harm-salience which may be moderated by anxiety.

H5. Frame-type and harm content will influence hypothetical moral judgments about the treatment of coronavirus patients.

H6. Moral judgments will predict behavioural intentions.

Method

Participants

The sample was 39.7% male, 1 MtF Female participant, and 1 person preferred not to disclose. The age of participants ranged between 18 and 74 years (Median=23, SD= 12.42). Five people did not disclose their age. A large proportion of the sample resided in Great Britain and Northern Ireland (47.7%). Minimum sample size was based on a-priori power calculations using G*Power ($\alpha = 0.05$, $\beta = 0.8$), estimating medium effect sizes (d=0.5, df=1) resulting in 128 participants, 32 per group. We exceeded this slightly (n=151) and group sizes were nearly identical. Participants were required to be aged 18+ but no other specific inclusion criteria needed to be met. During the time of the study, from mid-March through to early April 2020, 92% of the participants were in countries that were beginning or had already begun national lockdown. The remaining sample were either following localised or national recommendations, or in localised lockdown (Dunford et al., 2020).

Design

A 2x2 between participant's design investigated the impact of harm-content (high vs low) and frame-type (episodic vs thematic) of COVID-19 media narratives. Group sizes for each of the four conditions were roughly equal: high harm-content episodic group (n=36), low harm-content episodic group (n=38), high harm-content thematic group (n=39) and low

harm-content thematic group (n=38). There was no significant difference in the distribution of healthcare workers (F (3, 147) =.358, p=.783) or keyworkers (F (3, 147) =.608, p=.611), between group conditions. Each participant viewed one news article. Dependent variables included: anxiety, subjective harm-salience (SHS), moral judgments and behavioural intentions related to coronavirus. Interoceptive sensibility was a within-participants independent variable. Age, sex and baseline individual differences in people's experiences and perceptions towards the pandemic were control variables.

Materials

News articles

All articles were equal in word count; did not refer to any geographical location; and provided the same message at the end: *The message from government health officials remains the same:* ***Stay home. Protect our health service. Save lives***. All articles were based on real-news stories, but fictitious in nature (see supplementary material). Fictitious names used within the episodic articles were the same, and also for the thematic articles. No images were used. The articles were pre-validated by a group of five people who read the articles and rated whether the harm-salience of each article and whether the articles were more thematic or episodic in nature, which was followed up with a brief telephone discussion. The episodic articles focused on the individual stories of several people. The high harm-content episodic article depicted a distressing story about a family who had lost their son/sibling very quickly to the virus, whereas the low harm-content episodic article depicted several shopkeepers and pharmacists who were discussing their fears of catching the virus while at work. Quotes and personal experiences of individuals were the focus of these articles. The high harm-content thematic article discussed the wider impact and suffering of people within the health service dealing with an overwhelming amount of very ill patients. The low harm-content thematic article discussed the wider impact of government failings to roll out mass testing for coronavirus. Thematic articles did not focus on individual stories, although used quotes from 'experts' in the field providing comment on the wider story. The fundamental differences between high and low harm-content articles were the amount of harm-related language and descriptions used. More direct associations of harm and suffering of individuals/families (episodic) and larger bodies of people (thematic) were made in the high harm-content articles.

Moral dilemma stimuli

The moral dilemma task involved six situations directly related to treatment for coronavirus patients in hospital (see supplementary materials). Three of the dilemmas required participants to adopt the role of self, asking people to consider a situation involving a friend or relative (Personal). The other three dilemmas required participants to adopt the role of a healthcare worker involving patients with coronavirus (Impersonal). Presentation order was as follows: impersonal-personal-impersonal-personal-impersonal-personal. The moral dilemmas were based on a recent paper (Emanuel et al., 2020) outlining moral guidelines for healthcare workers in America during the coronavirus pandemic which instructed ethical treatment principles including: maximising outcomes offered by limited resources, equal treatment of people, prioritising people who were more ill or more at risk, and endorsing and prioritising instrumental value (e.g. supporting people such as healthcare workers who could be of help during the crisis), which were incorporated into the moral dilemmas. These principles are largely 'utilitarian' in nature in that they promote utility and maximising outcomes/advantage for the most people.

After reading each dilemma, participants were asked to judge whether a given action was morally appropriate or inappropriate e.g. '*Do you take the ventilator away from the one patient to share between the other two patients?'.* Response options included: 'Yes, this is appropriate' or 'No, this is not appropriate' (coded as Yes=1, No=0). Participants were then asked to rate how morally acceptable they judged the action that was propositioned in the question to be (1=Completely unacceptable, 7=Completely acceptable). For two of the personal moral dilemmas, participants were asked to judge whether an anti-utilitarian was morally appropriate e.g. '*Do you insist the ventilator remains with your friend?*'. These questions were reverse coded (Yes, this is appropriate =0, and No, this is inappropriate=1). Moral acceptability scores for these items were also reverse coded (i.e. Completely acceptable=1, Completely unacceptable =7) to inversely reflect judgments about the inappropriateness of anti-utilitarian actions. Mean moral appropriateness and moral acceptability scores were calculated for the categories of 'Impersonal' and 'Personal' dilemmas.

Behavioural intentions

After reading the article, participants were asked to rate how likely it was that they would carry out certain behaviours over the coming weeks (1=Unlikely, 7=Very likely). Behaviours included: 1) cutting out inessential shopping trips, 2) being more cautious around social distancing protocols 3) avoid going outdoors for walks/runs/cycles at 'busier' times 4) contact a friend/relative who might need support 5) offer to volunteer COVID-19 effort (where possible) 6) keep up to date with news surrounding coronavirus 7) purchase more food/cleaning supplies/medical supplies than needed 8) Go against government protocol to meet a friend/relative

Subjective harm-salience

At the end of the experiment participants were asked: *'In the article you read, how striking (or salient) was the physical/mental harm and suffering of individuals or groups of people dealing with Covid-19?'*. Likert scale responses included 1=Not very striking/salient and 7=Very striking/salient.

Anxiety

State anxiety was calculated using the State Anxiety Scale (STAI, Spielberger & Gorsuch, 1983). The scale includes positively and negatively coded items. Participants were asked to indicate their agreement (Not at all/Somewhat/Moderately so/Very much so), with twenty different statements e.g. 'I feel calm', 'I feel tense', 'I feel at ease'.

Interoceptive sensibility

People's awareness and response to both pleasant and unpleasant bodily sensations was measured using the 'noticing', 'not distracting' and 'not worrying' subscales of the Multidimensional Assessment of Interoceptive Awareness scale version 2 (Mehling et al., 2018). Participants were asked to indicate how often each item statement applied to them generally in daily life (0=Never, 5=Always). The noticing subscale consists of 4 items e.g. '*When I am tense, I notice where the tension is located in my body*'; the not-distracting subscale consists of 6 items which were reverse coded e.g. '*I distract myself from sensations of discomfort*'; and the not-worrying subscale consists of 5 items, 3 of which were reverse coded e.g. '*I can notice an unpleasant body sensation without worrying about it*'.

Individual differences

We used items from a recent pre-print investigating individual differences in people's responses to COVID-19 (Everett et al., 2020) that capture a number of baseline factors that could contribute to people's perceptions and reactions to news coverage about COVID-19. Items included: 1) *How much of a threat do you think COVID-19 (coronavirus) is?* (1 = not a threat at all; 7= extremely threatening) 2) *How likely do you think it is that you yourself will die as a result of the COVID-19 pandemic?* (1 = extremely unlikely, 7 = extremely likely); 3) *How likely do you think it is that someone you know will die as a result of the COVID-19 pandemic?* (1 = extremely unlikely, 7 = extremely likely); 4) *How unpleasant would it be for you personally to stay at home and avoid social contact for the next 2 weeks?* (1 = not at all unpleasant; 7 = extremely pleasant); 5) *Over the past week, how often have you sought out information about the COVID-19 pandemic* (e.g. news articles, video, social media posts, etc.)? (1 = rarely, 4 = occasionally, 7 = frequently); 6) *Are there currently a lot of cases of COVID-19 in the city where you live?* (yes, no, not sure); 7) *Has anyone you know personally tested positive for coronavirus?* (yes, no, not sure). Items 6 and 7 were coded as Yes=2, Unsure=1, and No=0.

Procedure

Following approval from University of Bath Psychology Ethics Committee, 151 participants were recruited online via the research platform Prolific, University of Bath research participation pool and departmental email lists. Participants recruited via Prolific were awarded monetary compensation for their time in accordance with platform guidelines. Students accessing the experiment via the University participation pool were awarded credits as part of the experimental hours scheme. All other participants were able to enter into a prize draw to win 1 of 4 £20 Amazon vouchers. The experiment was developed in Qualtrics online experimental software and accessed via an anonymous link. Prospective participants were explicitly warned in research advertisements and information sheets about the potential upsetting nature of the experiment. Participants viewed an information sheet and consent form and encouraged to contact the experimenter with any questions or concerns about taking part. The whole experiment took 20-25 minutes. Demographic information, baseline individual differences measures and the interoceptive sensibility measures were completed first. Participants were then presented with 1 of 4 news articles and reminded throughout that they could close their browser window at any time to end the experiment. Presentation of

news articles was randomised, to ensure equal group sizes. There was no time limit for participants to read the news article. Participants then completed the state anxiety scale. Instructions for the moral dilemma task were then presented. There were six moral dilemma stories, viewed in the same order. The two moral judgment questions were presented after each dilemma. There was a timing function set for the moral dilemmas, so that 7-10 seconds had to pass before participants could advance to the questions depending on the amount of text. After 50 seconds participants were automatically advanced to the moral judgment questions. The final section required participants to rate their intentions to carry out certain behaviours over the next few weeks. Participants were then asked to judge how salient they perceived the physical or mental harm of people in the news article to be. A debrief sheet followed providing further details about the research and information and links for people to seek further support if needed. Other links to the World Health Organisation regarding coronavirus were also provided, as well as YouTube videos which aimed to provide a positive or restorative experience for participants.

Results

Individual differences in experiences and perceptions of coronavirus

Around 40% of the sample were living in a place with many cases of coronavirus and also knew someone who had tested positive for coronavirus (Table 1), indicating a considerable amount of people had been personally impacted by the pandemic. A principal component analysis (PCA) was carried out on the remaining five individual differences questions to see whether they could be reduced to a single dimension. Four factors were retained which cumulatively explained 91.1% of the total variance with eigenvalues of 1.934, 0.854, 0.692, and 1.076. The two component items combined (item 2 and 3) both represented a sense of threat regarding the deadliness of COVID-19 to the self and others. A mean score of these items was relabelled ‘Perceived deadliness of virus’. The remaining items were retained as independent scores.

Table 1

Percentage of sample who were aware of lots of cases of coronavirus in the place that they lived, and percentage of sample who knew someone who had tested positively for coronavirus

	Yes	Unsure	No
Many cases nearby	39.7	14.6	45.7
Tested positive for virus	39.7	6.6	53.6

Behavioural intentions

A principal component analysis (PCA) was also carried out to investigate whether the behavioural items could be grouped within distinct dimensions. Three factors were retained (Table 2). These factors cumulatively explained, 30.15%, 46.57% and 61.19% of the total variance with Eigenvalues of 2.248, 1.477 and 1.171 respectively. Items for each factor were combined into a composite mean score. Factor 1 was labelled 'Preventative behaviours', as all items reflected intentions to reduce acts that could increase the catching or spread of the virus. Factor 2 was labelled 'Prosocial behaviours' as these items represented a sense of carrying out acts which had positive outcomes for others, including taking responsibility for keeping informed via the news. Factor 3 was labelled 'Antisocial behaviours', as these final two items corresponded with actively selfish behaviours that could have negative outcomes for others.

Table 2

*Varimax Rotated component Matrix for Principal Component analysis. Values refer to factor loadings. Items in **bold** face retained for that factor.*

Behaviour	Factor 1	Factor 2	Factor 3
	Preventative	Prosocial	Anti-social
1. Reduce shopping trips	**.876**	.085	-.077
2. Caution social distancing	**.783**	.312	-.017
3. Avoid outdoors	**.759**	-.043	.267
4. Contact friend/relative	.133	**.825**	-.067
5. Volunteer	-.033	**.619**	.344
6. News updates	.086	**.546**	-.154
7. Bulk-buying	.196	-.079	**.836**
8. Break protocol	-.477	.034	**.498**

H1. Frame-type, harm-content and anxiety

A two-way analysis of variance (ANOVA) was conducted to test the effects of frame-type (episodic versus thematic) and harm- content (high harm- content versus low harm-content) on anxiety (see Table 3). Contrary to H1, there was no significant interaction between harm- content and frame-type for anxiety scores, F (1, 147) =.586, p=.445, partial η2=.04. Simple main effects for harm- content F (1, 147) =.509, p=.477, partial η2=.003, and frame-type, F (1, 147) =1.711, p=.193, partial η2=.012, were also non-significant. This suggests that frame-type and harm-content of articles people read were not responsible for the differences in anxiety people reported after reading the articles.

Table 3

Mean and standard deviation anxiety scores for each group condition (Harm content=HC)

	Article group			
	Episodic		Thematic	
Anxiety	Low HC	High HC	Low HC	High HC
Mean	45.63	48.86	44.45	44.33
SD	14.37	10.42	13.71	14.53

H2. Interoceptive sensibility and anxiety

A three-step hierarchical regression was performed to test whether a tendency to notice, focus on, or worry about internal sensations predicted feelings of anxiety when controlling for group condition and key individual differences. Subjective harm-salience (SHS) was also included as a predictor, to test its association with anxiety. Predictors entered at step-one were age, sex, threat of COVID-19, subjective deadlines of COVID-19, unpleasantness of staying at home, many cases of COVID-19 in area, known someone who tested positive with coronavirus, and seeking information about coronavirus. Frame-type and harm-content condition were entered at step-two. SHS scores and interoception subscales ‘noticing’, ‘not distracting’ and ‘not worrying’ were inputted at step-three.

Model 1 significantly predicted anxiety, R^2 *chng*=.216, *F* (8, 137) =4.179, *p*<.0005. Specifically, age (*b*=-.247, *p*=.004), sex (*b*=.178, *p*=.026), and finding it unpleasant to be at home during lockdown (*b*=.172, *p*=.029) all significantly predicted anxiety in Model 1-3, with younger people and women reporting higher levels of anxiety. The addition of frame-type and harm-content in Model 2 did not significantly improve the model. In Model 3, we found that including SHS (*b*=.205, *p*=.020) and the interoception 'not worrying' subscale (*b*=-.169, *p*=.039) led to a statistically significant increase (*F change*= 3.144, *p*=.017) of variance explained, R^2 *chng*=.067, *F* (14, 131) =3.979, *p*<.0005. Therefore, in support of H2 people's appraisals of harm-salience and a tendency to worry about bodily sensations significantly increased feelings of anxiety when controlling for group condition and key individual differences.

H3. Frame-type, harm-content and subjective harm salience

A two-way Analysis of Covariance (ANCOVA) was conducted to assess the impact of frame-type and harm-content on subjective appraisals of harm-salience (SHS). SHS was the dependent variable (Table 4), with two between subjects' factors: frame-type (episodic versus thematic) and harm-content (high versus low). The three interoception subscales were included as covariates to explore the influence of frame-type and harm-content when controlling for interoceptive tendencies. The ANCOVA revealed a main effect of harm-content, *F* (1, 144) =26.32, *p*<.001, partial η2=.155, a main effect of frame-type, *F* (1, 144) =7.457, *p*=.007, partial η2=.049, and a significant interaction between frame type and harm-content on SHS ratings, *F* (1, 144) =5.457, *p*=.021, partial η2=.037. An analysis of simple main effects of frame-type and harm-content was performed with statistical significance receiving a Bonferroni adjustment. Bootstrapping (x1000) was applied to confidence intervals and tests of significance of post hoc tests. We found SHS was significantly higher for the episodic low harm-content article compared to thematic low harm-content article, *F* (1, 144) =12.821, *p*=.001, partial η2=.082, 95% CI [-1.701 to -.450]. Although SHS was marginally higher for the episodic high harm content article than the thematic high harm content article, this was not statistically significant, *F* (1, 144) =0.95, *p*=.740, partial η2=.001. Therefore, in partial support of H3, episodic frames strengthened appraisals of harm-salience, but only for stories not containing very harm-salient content. For both thematic and episodic stories that have more harm-salient content, people perceived the harm and suffering of individuals versus larger groups/organisations in society as similarly striking.

For the interoception covariates, the 'noticing', F (1, 144) =2.772, p=.098, partial η2=.019, and 'not distracting' subscales, F (1, 144) =1.443, p=.232, partial η2=.010, did not significantly explain variance in SHS ratings. However, the 'not worrying' subscale was significantly related (r= -.133) to SHS, F (1, 144) =4.360, p=.039, partial η2=.029. This suggests people who were more prone to worrying about painful/unpleasant bodily sensations gave higher subjective-harm salience ratings when controlling for frame-type and harm-content.

Table 4

Means, Standard Deviations and Standard errors for Subjective harm-salience ratings (SHS) for the Four Article Groups (HC = Harm-content)

	Article group			
	Episodic		Thematic	
SHS	Low HC	High HC	Low HC	High HC
M	4.842	5.361	3.763	5.230
(SD)	1.424	1.198	1.459	1.111
(SE)	.241	.194	.235	.179

H4. Frame type, harm-content, subjective harm salience and behavioural intentions

We investigated whether harm-content and frame-type indirectly influenced participants preventative, prosocial and antisocial behaviours by shaping their appraisals of harm salience. Furthermore, because subjective harm salience (SHS) predicted feelings of anxiety in H2, and anxiety has been shown to be an important factor determining preventative behaviours in previous pandemics, we tested whether the relationship between subjective harm-salience (SHS) and behaviour was modulated by anxiety. We used PROCESS Model 21 (Hayes, 2018a; see figure 1) to explore the influence of harm-content and frame-type for each of the behaviour measures, resulting in two regression analyses for the three dependent variables a) Preventative b) Prosocial and c) Antisocial behaviours. For each dependent variable, harm-content was inputted as the independent variable and frame-type as the

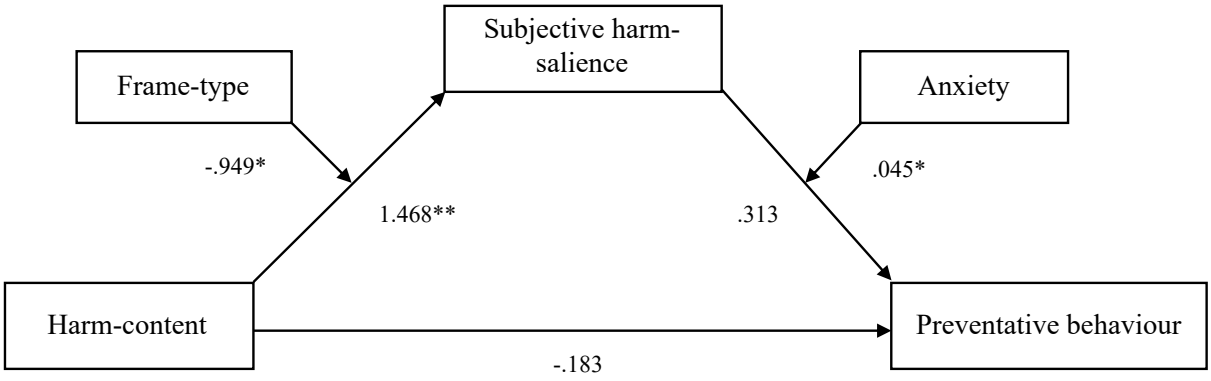
moderator between harm-content and SHS. Anxiety was inputted as a moderator between SHS and behaviour. For all models, we found a greater amount of harm-content (b=1.468, p=<.0001) and the use of episodic frames (b=1.079, p=.0004) strengthened subjective harm-salience (SHS), R^2= *.1915, F* (3, 147) =11.606, p<.0001.

Preventative behaviours

We found frame-type moderated the effect of harm-content on SHS, with harm-content only increasing SHS in the thematic style articles, R^2 *chng= .028, F* (1, 147) =4.966, p=.027, 95% CI [.8788 to 2.056]. The interaction between SHS and anxiety significantly predicted preventative behaviours, b=.0114, p=.045, R^2 *= .025, F* (1, 146) =4.068, p=.045, 95% CI [.0002 to .0225], indicating a moderation effect of anxiety (Figure 1). The Johnson-Neyman (JN) technique was used to explore this moderation effect by identifying regions of significance. We found SHS increased preventative behaviours but only for people scoring above 39.08 points on the anxiety scale, 69.5% of the sample. We also found that the effect of harm-content of thematic articles through the mediating factor of SHS indirectly increased preventative behaviours, but only for people scoring at the mean and 1 standard deviation above the mean on the anxiety scale. Anxiety did not moderate the overall indirect effect of harm content on preventative behaviours, 95% CI [-.0279 to .0003], suggesting the modulating influence of anxiety was isolated to the pathway between SHS and preventative behaviours (Hayes, 2015, 2018b). In sum, a greater amount of harm content in the thematic articles strengthened appraisals of harm-salience, which, given a sufficient level of anxiety, influenced them to commit to preventative behaviours.

Figure 1.

Statistical diagram of PROCESS Model 21 (Hayes, 2018) for effect of harm-content on subjective harm salience (SHS) and preventative behaviour



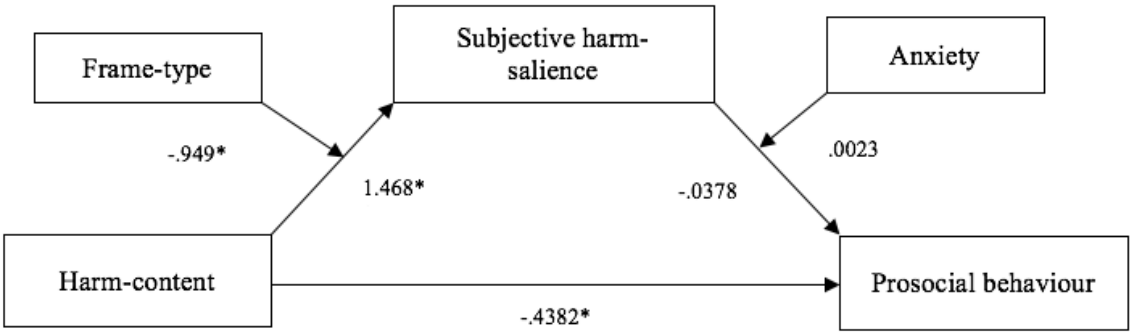
Note. Harm-content (high= 1, low=0) and frame-type (episodic=1, thematic=0). Effect of subjective harm-salience on preventative behaviour is moderated by anxiety. Values represent beta coefficients for each pathway of the model, from left to right: interaction frame-type*harm-content on SHS; effect harm-content on SHS; direct effect of harm-content; effect of SHS on preventative behaviour; interaction SHS*anxiety on preventative behaviour (*p<.05, **p<.00001).

Prosocial behaviours

Again, we used PROCESS Model 21 (Hayes, 2018a) to investigate whether frame-type and harm-content indirectly influenced prosocial behaviour through the mediating factor of subjective harm-salience (SHS). We found a direct effect of harm-content on prosocial behaviours, t=.5253, p=.0226, 95% CI [-.8139 to -.0625]. Specifically, people who read the high harm-content articles were less likely to commit to prosocial behaviours, such as calling a relative or keeping up to date with the news and this was independent from their appraisals of harm-salience and anxiety. Confidence intervals for the indirect effects of harm-content on prosocial behaviours through SHS included zero, indicating no indirect effect of harm-content on prosocial behaviours through SHS, R^2= *.036, F* (4, 146) =1.385, p=.242 (Figure 2).

Figure 2.

Statistical diagram of PROCESS Model 21 (Hayes, 2018) for effect of harm-content on prosocial behaviours



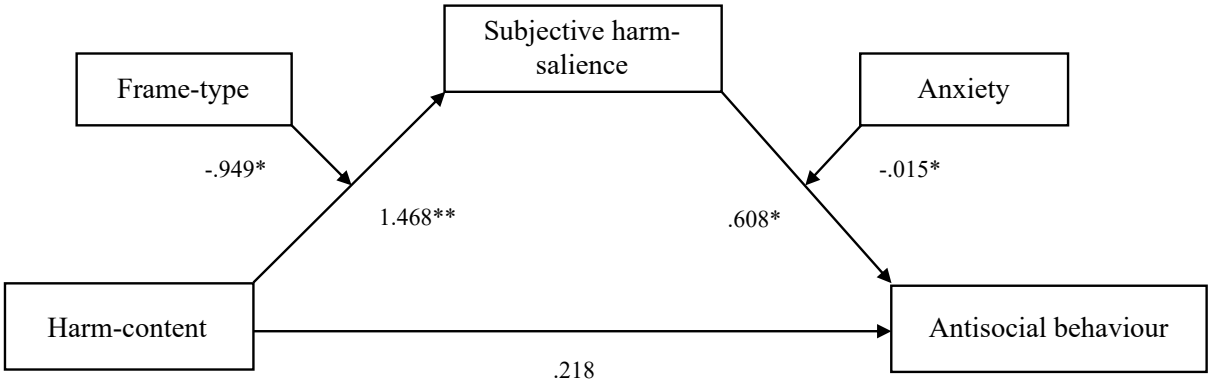
Note. Harm-content (high=1, low=0) and frame-type (episodic =1, thematic =0).Values represent beta coefficients for each pathway of the model from left to right: interaction frame-type*harm-content on SHS; effect harm-content on SHS; direct effect of harm-content; effect of SHS on prosocial behaviour; interaction SHS*anxiety on prosocial behaviour (*p<.05, **p<.0005).

Antisocial behaviours

Using PROCESS Model 21 (Hayes, 2018a) we investigated whether frame-type and harm-content indirectly influenced antisocial behaviour. We found subjective harm salience (SHS) significantly predicted antisocial behaviours, b=.608, p=.025, R^2= *.056, F* (4, 146) =2.173, p=.075, but again, this was moderated by anxiety, R^2 *chng= .027, F* (1, 146) =4.214, p=.042, 95% CI [-.0228 to -.0004]. Using the JN technique, we found SHS only predicted antisocial behaviour for people scoring below 36.29 on the anxiety measure, 28.48% of the sample. The overall indirect effect of harm-content on antisocial behaviour was also moderated by anxiety 95% CI [.0000 to .0286]. Therefore, thematic articles higher in harm-content increased the likelihood that people would carry out antisocial behaviours by enhancing appraisals of harm-salience, but only for people experiencing relatively low levels of anxiety after reading the article (Figure 3). This suggests a moderate level of emotional distress may be important to deter people from carrying out more selfish or antisocial behaviours during the pandemic.

Figure 3.

Statistical diagram of PROCESS Model 21 (Hayes, 2018a) for effect of harm-content on subjective harm salience (SHS) and antisocial behaviour



Note. Harm-content (high=1, low=0) and frame-type (episodic =1, thematic =0). Indirect effect of harm-content moderated by anxiety. Values represent beta coefficients for each pathway of the model from left to right: interaction frame-type*harm-content on SHS; effect harm-content on SHS; direct effect of frame-type; effect of SHS on antisocial behaviour; interaction SHS*anxiety on antisocial behaviour (*p<.05, **p<.0005).

H5. Frame-type, harm-content and moral judgments

We investigated whether the harm-content and frame-type of articles influenced people's hypothetical moral judgments regarding the treatment of coronavirus patients. Moral appropriateness judgments included both impersonal (adopting role of healthcare worker) and personal (adopting role of self) dilemmas. Participants subsequently rated how morally acceptable they found the proposed actions within the moral dilemma to be. We conducted a repeated-measures ANOVA to test the effect of dilemma type; 'Personal' versus 'Impersonal', on people's moral appropriateness and moral acceptability judgments. Harm-content and frame-type were included as between-participant factors. A Pearson's bivariate correlation showed none of the demographic, individual differences or key independent variables correlated with moral judgments so were not included in the model.

The ANOVA revealed the average moral appropriateness scores for the 'Personal' dilemmas were statistically lower than for the 'Impersonal' dilemmas, F (1,147) = 19.644, p<.001, partial η2=118. This shows that people were more likely to condone utilitarian action

when in the role of a healthcare worker, compared to dilemmas where participants made decisions that could impact their friends or relatives. There was no interaction between dilemma type and harm-content, F (1,147) = .258, p=.612, partial η2=.002, or frame-type, F (1,147) = <.001, p=.998, partial η2<.001, or between all three factors, F (1,147) = .097, p=.756, partial η2=.001. Therefore, across all groups people typically judged utilitarian action in the impersonal dilemmas to be more morally appropriate compared to the personal dilemmas. A further repeated measures ANOVA was conducted for moral acceptability judgments. This time, no effect of dilemma type was found, F (1,147) = .019, p=.890, partial η2<.001. Therefore, people rated the moral acceptability of utilitarian action similarly in both personal and impersonal dilemmas. There was no interaction between dilemma type and harm-content, F (1,147) = 1.503, p=.222, partial η2=.010 or frame-type, F (1,147) = 1.322, p=.252, partial η2=.009 , or between all three factors, F (1,147) = .140, p=.709, partial η2=.001. Therefore, people typically judged the acceptability of utilitarian actions as more similar, regardless of the personal connection to the people in the story. All findings were sustained when excluding all health workers from the analyses. Overall, the harm-content and frame-type of coronavirus news coverage did not affect people's moral judgments directly related to COVID-19 patients.

H6. Moral judgments and behavioural intentions

We used a multivariate regression approach to investigate whether hypothetical moral judgments related to the coronavirus crisis predicted people's behavioural intentions associated with COVID-19. Prosocial, antisocial, and preventative behaviours were inputted as the dependent variables. 'Personal' moral appropriateness scores, 'Impersonal' moral appropriateness scores, 'Personal' moral acceptability ratings and 'Impersonal' moral acceptability ratings were entered as predictors. We found a main effect of Personal moral acceptability scores, F (3, 144) =3.326, p=.021, partial η2=.065. There were no main effects of any other predictors. Parameter estimates providing the unstandardized regression coefficients, indicated that higher moral acceptability ratings in personal moral dilemmas was positively associated with prosocial behaviour during the pandemic, B=.311, p=.006, 95% CI [.091 to .531]. Therefore, people who saw it as more acceptable to support utilitarian action in their moral judgments, even if that means sacrificing the care of a loved one, may be more likely to behave in ways that benefit others during the pandemic.

Discussion

The media is a powerful source of both information and influence during the coronavirus pandemic as we look to news coverage to communicate guidance, alarm and reassurance about the evolving health crisis. Here, we explored the influence of a fictitious news article on people's feelings of anxiety, behavioural intentions associated with the pandemic and moral judgments associated with the treatment of coronavirus patients. Contrary to H1, we did not find that episodic frames or articles higher in harm content increased readers anxiety. Instead, in support of H2 we found a tendency to worry about painful or unpleasant bodily sensations, as well as people's subjective appraisals of harm-salience predicted feelings of anxiety. In H3 we found an interaction between harm-content and frame-type influenced subjective appraisals of harm salience; specifically, a greater amount of harm-content in the thematic articles, and the use of an episodic frame in low harm-content articles strengthened appraisals of harm-salience. Interoceptive sensibility (not worrying) also independently predicted stronger harm-salience appraisals. In H4, we discovered that higher harm-content in thematic articles, indirectly increased the likelihood people would carry out preventative (e.g. social-distancing) and antisocial (e.g. bulk-buying) behaviours, by enhancing appraisals of harm salience. These effects, however, were conditional upon a certain level of anxiety experienced after reading the articles. A higher level of harm-content also directly decreased the likelihood people would carry out prosocial behaviours. We did not find any effects of frame-type or harm content on people's moral judgments (H5), and people's moral judgments about the treatment of coronavirus patients did not predict their real-world behavioural intentions during the pandemic (H6).

Effects of harm-content and frame-type on harm salience appraisals

Previous work has shown that episodic frames, explicitly identifying victims are more effective at evoking an emotional response from readers (e.g. Aarøe, 2011; Kogut & Ritov, 2005b; Small & Loewenstein, 2003). Our findings suggest this 'singularity effect' did not apply to appraisals of harm-salience, as a thematic narrative that was high in harm-content was shown to be as powerful for enhancing readers' appraisals of harm-salience as an episodic narrative describing a family's loss of a son to coronavirus. It is possible that the language and metaphors used to describe harm in this article had anthropomorphic qualities, which may have generated a more humanised representation of the health service, in a similar way to narratives focusing on the emotional experiences of an individual person.

Alternatively, the high harm-content articles may have been more effective at amplifying risk or threat (Kasperson et al., 2016) associated with coronavirus, due to the balance of information regarding the harmful consequences of COVID-19 (Chang, 2012; Klemm et al., 2016). If participants considered the self-relevant consequences of the harm and suffering of others when forming their appraisals, these evaluations may represent some form of self-oriented harm-salience which was projected onto those within the article. Further work is needed to understand whether appraisals regarding the harm and suffering of others depicted in media coverage, correlates with self-oriented evaluations of harm, threat or risk discussed in the wider health communication literature (e.g. Klemm et al., 2016).

Interoceptive sensibility, anxiety and harm-salience

Media coverage has the potential to exacerbate public anxieties of external threats (e.g. Bodas et al., 2015; Tabri et al., 2020). Here we have found that individual differences in interoception may also provide a vulnerability to experience negative emotional responses to distressing media coverage. People with a tendency to worry more about painful or unpleasant interoceptive sensations gave stronger appraisals of harm-salience overall and were more likely to report feeling anxious after reading the articles. Previous research has shown that the interoception 'not worrying' dimension is associated with trait anxiety and susceptibility to emotionally laden stimuli (Calì et al., 2015; Mehling et al., 2018). This may explain why people high in interoceptive sensibility found the harm and suffering of people in the news articles to be more salient. Importantly, the average level of anxiety across all groups (scores between 44.33-48.86) was considerably higher than healthy population norms (e.g. Julian, 2011; Knight, Waal Manning, & Spears, 1983), which is perhaps not surprising in the context of a global pandemic. Future research incorporating pre and post-treatment anxiety measures could determine whether people with higher levels of interoceptive sensibility experience greater increases in anxiety than others, after exposure to pandemic news coverage.

The indirect effect of frame-type and harm-content on behavioural intentions

A greater amount of harm-content in the thematic articles, and the use of an episodic frame in low harm-content articles indirectly boosted intentions to carry out preventative behaviours by enhancing perceptions of harm-salience. This may be because episodic frames have been shown to encourage individual action by enhancing the saliency of individual

responsibility compared to thematic frames (Hart, 2011; Iyengar, 1994). Yet, the fact that episodic frames also indirectly increased antisocial behavioural intentions challenges the notion that episodic frames were more effective at inducing a sense of social responsibility.

Anxiety and behavioural intentions

A sufficiently high level of anxiety after reading coronavirus news coverage was a key moderating factor determining whether appraisals of harm-salience translated into preventative behaviours, and is consistent with epidemiological research (Bults et al., 2011; Jones & Salathé, 2009; Lau et al., 2003; Leung et al., 2005). Interestingly, the indirect effects of harm-content and frame-type for predicting antisocial behaviours was only present for people experiencing low levels of anxiety. This novel finding indicates a dampened emotional response after reading emotive coronavirus articles may lead to more selfish behaviours that could have consequences for viral spread. Anxiety did not modulate the influence of subjective harm salience on prosocial behaviours, suggesting anxiety did not facilitate behaviours benefitting the interests of others. Therefore, as in Hart (2011) we suggest that coronavirus articles that evoked a sufficient level of emotional distress, promoted egoistically motivated preventative behaviours in an attempt to reduce emotional distress. Finally, news articles higher in harm-content directly reduced people's intention to carry out prosocial behaviours including, contacting a friend or relative who might need support and keeping informed about the pandemic via the news, which is problematic as keeping alert to media information appears to play an important role in encouraging preventative behaviours (Bults et al., 2011). Striking the correct balance of emotional arousal has been shown to be important for engagement with evocative news coverage (Vettehen, Nuijten, & Peeters, 2008) which may explain this finding.

Influence of coronavirus news coverage on moral judgments

Readers' hypothetical moral judgments about the treatment of coronavirus patients were less impacted by coronavirus media coverage. Although people typically judged anti-utilitarian acts in the personal moral dilemmas (adopting the role of self) to be more morally appropriate than in the impersonal moral dilemmas (adopting role of healthcare worker), these effects were not influenced by either the frame-type or harm-content of the articles. It is not surprising that people judged anti-utilitarian or more 'self-serving' actions to be more morally appropriate for the dilemmas involving friends and family, as self-interested

motivations are likely to bias our moral judgments to be less utilitarian (Thomas, Croft, & Tranel, 2012). The finding that moral acceptability judgments did not show the same pattern of responses between personal and impersonal moral dilemmas is consistent with psychological (Tassy et al., 2013) and neuropsychological studies (Berthoz et al., 2006) that have found inconsistencies between appropriateness judgments (egocentric) and acceptability (allocentric) judgments.

Dissociation between moral judgments and behaviour

People's moral preferences associated with the treatment of coronavirus patients provided little indication of their behaviours during the pandemic. This is consistent with virtual reality research that has found the hypothetical moral judgments people make, and the moral behaviours they carry out in more realistic environments, are often contradictory (e.g. Francis et al., 2016; Francis et al., 2017). We did find people who judged it to be more morally acceptable to carry out utilitarian action in the personal moral dilemmas, were more likely to self-report the intention to carry out prosocial behaviours in the coming weeks, such as contacting a loved one or volunteering. Therefore, people who were more willing to hypothetically reject superior treatment for a friend or relative [if superior treatment would result in poorer outcomes for other people] were more likely to carry out behaviours that could have positive outcomes for others during the pandemic. It is possible that these people had an underlying altruistic motivation to increase welfare to others (Batson, Fultz, & Schoenrade, 1987). Interestingly, people's moral judgments did not correlate with their appraisals of harm-salience or anxiety, unlike the behavioural measures. This suggests that the harm-salience of coronavirus news coverage was impactful for shaping individual behavioural action but did not sway their underlying moral perspectives.

Limitations

We acknowledge several limitations of this research. This study was carried out during a unique period of time, when coronavirus was beginning to spread rapidly around the world and many countries began to go into lockdown for the first time. The fictitious articles created were based on real UK news media, potentially providing a UK-biased perspective on issues relating to coronavirus. The findings should be considered within this specific context of public discourse in the media, as the public health challenges associated with coronavirus

have been unprecedented and continue to develop. Further work in other epidemics or pandemics is needed to assess the transferability of findings in other media contexts.

Secondly, our measure of harm-salience is novel and based on a single item question. How harm-salience correlates with other variables such as self-oriented emotional distress associated with empathy (Hart, 2011; Sarlo et al., 2014), and risk perception (Klemm et al., 2016), warrants further exploration. In addition, although each article was evaluated as representing categories of high and low harm-content and episodic versus thematic frame-types, each article discussed different coronavirus 'topics' which potentially influenced participants in ways we did not capture. Stricter standardisation of media coverage in future between-subjects designs would control for content-based confounding variables. Further studies would also benefit from including an attention check to be confident that participants were engaged with the online tasks. Finally, the behavioural measures we used reflected participants 'intention' to carry out preventative, prosocial and antisocial behaviours, and do not necessarily guarantee real-world behaviours. Longitudinal designs collecting health-relevant behavioural data during a pandemic would provide invaluable insights into the ecological validity of these findings.

Media framing as a public health strategy

Media frames that were more effective at increasing the saliency of harm and suffering of people impacted by coronavirus influenced how anxious people felt and the socially relevant behaviours they intended to carry out during the pandemic. Highly harm-salient media narratives may represent a double-edged sword which could simultaneously encourage people to minimise behaviours that increase the spread of the virus, whilst also motivate them to acquire more essential resources than they need or breach public health protocols. Deliberately using framing devices as a strategy to influence public health raises fundamental ethical questions. Increasing the salience of certain issues in the minds of audiences makes readers vulnerable to elite manipulation and may undermine citizen competence (Druckman, 2001). However, some evidence suggests that people interpret and respond to media frames through the lens of their existing predispositions and perspectives (Gross & D'Ambrosio, 2004). Our findings somewhat support this perspective, as it was people's personal appraisals of harm-salience, as opposed to the news coverage manipulations per se, that were influential for emotional and behavioural responses to coronavirus media coverage. Secondly, an important job of public health communications

during a pandemic is to effectively support the public to manage difficult emotions (Lee & Basnyat, 2013), which could be hampered by sensationalist news pieces that amplify fear (Klemm et al., 2016). However, others have suggested that more sensational news coverage can be socially functional for drawing people's attention towards what acts, or events are deemed to be morally acceptable in society (Stevens, 1985). Quantifying the balance of 'emotional' versus factual coronavirus media coverage demands a retrospective analysis of the interactions between content, tone, and frame-types (Klemm et al., 2016), as well as the perspectives of the receiver (Gross & D'Ambrosio, 2004). Considering the emerging nature and scale of the coronavirus pandemic, judging whether dramatising media coverage about coronavirus is justified may only be possible following a review of the global impact of the virus retrospectively.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

H.B. conceived the study idea, carried out the research, conducted data analysis and paper writing. D.S.F. and M.P. supervised the entire project, piloted the research and provided feedback and ideas on the study design, data analysis and written paper. H.F. provided support with data analysis and feedback on written paper.

Acknowledgments

We thank members of the CREATE lab and Crossmodal Cognition lab for feedback on methodology. We thank Dr David Green and Johanna Meyer for useful discussions.

References

Aarøe, L. (2011). Investigating frame strength: The case of episodic and thematic frames. *Political Communication*, *28*(2), 207–226. https://doi.org/10.1080/10584609.2011.568041

Barlow, D. H. (1988). *Anxiety and its disorders*. New York: Guilford Press.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. *Journal of Personality*, *55*, 19–40.

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage*, *31*(2), 945–950. https://doi.org/10.1016/j.neuroimage.2005.12.039

Bodas, M., Siman-Tov, M., Peleg, K., & Solomon, Z. (2015). Anxiety-inducing media: The effect of constant news broadcasting on the well-being of Israeli television viewers. *Psychiatry (New York)*, *78*(3), 265–276. https://doi.org/10.1080/00332747.2015.1069658

Bolls, P. D., Lang, A., & Potter, R. F. (2001). The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial muscular Responses to Radio Advertisements. *Communication Research*, *28*(5), 627–651.

Bults, M., Beaujean, D. J. M. A., De Zwart, O., Kok, G., Van Empelen, P., Van Steenbergen, J. E., … Voeten, H. A. C. M. (2011). Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: Results of three consecutive online surveys. *BMC Public Health*, *11*, 1–13. https://doi.org/10.1186/1471-2458-11-2

Calì, G., Ambrosini, E., Picconi, L., Mehling, W. E., & Committeri, G. (2015). Investigating the relationship between interoceptive accuracy,interoceptive awareness,and emotional susceptibility. *Frontiers in Psychology*, *6*(August), 1–13. https://doi.org/10.3389/fpsyg.2015.01202

Chang, C. (2012). News Coverage of Health-Related Issues and Its Impacts on Perceptions: Taiwan as an Example. *Health Communication*, *27*(2), 111–123. https://doi.org/10.1080/10410236.2011.569004

Clark, D. M., Salkovskis, P. M., Breitholtz, E., Westling, B. E., Öst, L. G., Koehler, K. A., … Gelder, M. (1997). Misinterpretation of body sensations in panic disorder. *Journal of Consulting and Clinical Psychology*, *65*(2), 203–213. https://doi.org/10.1037/0022-006X.65.2.203

Craig, B. A. D. (2015). *How do you Feel?:An Interoceptive Moment with Your Neurobiological Self*. New York: Princeton University Press.

Deardorff, P. A., Kendall, P. C., Finch, J. A., & Sitarz, A. M. (1977). Empathy, locus of control and anxiety in college students. *Psychological Reports*, *40*, 1236–1238.

Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, *30*(1), 1–11. https://doi.org/10.1016/j.cpr.2009.08.008

Druckman, J. N. (2001). The Implications of Framing Effects for Citizen Competence. *Political Behavior*, *23*(3), 225–256. Retrieved from http://www.ingentaconnect.com/content/klu/pobe/2001/00000023/00000003/00369814

Dunford, D., Dale, B., Stylianou, N., Lowther, E., Ahmed, M., & de la Torre Arenas, I. (2020). Coronavirus: The world in lockdown in maps and charts. Retrieved from https://www.bbc.co.uk/news/world-52103747

Emanuel, E. J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., … Phillips, J. P. (2020). Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *The New England Journal of Medicine*, 1–7. https://doi.org/10.1056/NEJMsb2005114

Everett, J., Colombatto, C., Chituc, V., Brady, W. J., & Crockett, M. J. (2020). The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic. *PsyArXiv Preprints*, 1–23. Retrieved from https://psyarxiv.com/9yqs8

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Anderson, G. G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0164374

Gamson, W., & Modigliani, A. A. (1987). The Changing Culture of Affirmative Action. In R. D. Braungart (Ed.), *Research in political sociology* (pp. 3–137).

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: “Anterior insular cortex mediates bodily sensibility and social anxiety” by Terasawa et al.,(2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234. https://doi.org/10.1093/scan/nss140

Goldenberg, J. L., & Arndt, J. (2008). The Implications of Death for Health: A Terror Management Health Model for Behavioral Health Promotion. *Psychological Review*, *115*(4), 1032–1053. https://doi.org/10.1037/a0013326

Greenberg, J., Martens, A., Jonas, E., Eisenstadt, D., Pyszczynski, T., & Solomon, S. (2003). Psychological Defense in Anticipation of Anxiety: Eliminating the Potential for Anxiety Eliminates the Effect of Mortality Salience on Worldview Defense. *Psychological Science*, *14*(5), 516–519. https://doi.org/10.1111/1467-9280.03454

Gross, Kimberley. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, *29*(2), 169–192. https://doi.org/10.1111/j.1467-9221.2008.00622.x

Gross, Kimberly, & D’Ambrosio, L. (2004). Framing Emotional Response. *Political Psychology*, *25*(1), 1–29. https://doi.org/10.1111/j.1467-9221.2004.00354.x

Grover, K. W., & Miller, C. T. (2014). Effects of mortality salience and perceived vulnerability on HIV testing intentions and behaviour. *Psychology and Health*, *29*(4), 475–490. https://doi.org/10.1080/08870446.2013.866672

Hart, P. S. (2011). One or many? The influence of episodic and thematic climate change frames on policy preferences and individual behavior change. *Science Communication*, *33*(1), 28–51. https://doi.org/10.1177/1075547010366400

Hayes, A. F. (2015). An Index and Test of Linear Moderated Mediation. *Multivariate Behavioral Research*, *50*(1). https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018a). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd ed.). The Guildford Press: New York.

Hayes, A. F. (2018b). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Iyengar, S. (1994). *Is anyone responsible?:How television frames political issues.* University

of Chicago Press.

Iyengar, S., & Simon, A. (1993). News coverage of the Gulf crisis and public opinion: A study of agenda-setting, priming, and framing. *Communication Research*, *20*(3), 365–383.

Jones, J. H., & Salathé, M. (2009). Early assessment of anxiety and behavioral response to novel swine-origin influenza a(H1N1). *PLoS ONE*, *4*(12), 2–9. https://doi.org/10.1371/journal.pone.0008032

Julian, L. J. (2011). Measures of Anxiety. *Arthritis Care*, *63*(0 11), 1–11. https://doi.org/10.1002/acr.20561.Measures

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., … Ratick, S. (2016). The social amplification of risk: A conceptual framework. *The Perception of Risk*, *8*(2), 232–245.

Klemm, C., Das, E., & Hartmann, T. (2016). Swine flu and hype: A systematic review of media dramatization of the H1N1 influenza pandemic. *Journal of Risk Research*, *19*(1), 1–20. https://doi.org/10.1080/13669877.2014.923029

Knight, R. G., Waal Manning, H. J., & Spears, G. F. (1983). Some norms and reliability data for the State-Trait Anxiety Inventory and the Zung Self-Rating Depression scale. *British Journal of Clinical Psychology*, *22*(4), 245–249. https://doi.org/10.1111/j.2044-8260.1983.tb00610.x

Kogut, T., & Ritov, I. (2005a). The "identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/10.1002/bdm.492

Kogut, T., & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, *97*(2), 106–116. https://doi.org/10.1016/j.obhdp.2005.02.003

Kouchaki, M., & Desai, S. D. (2015). Anxious, threatened, and also unethical: How anxiety makes individuals feel threatened and commit unethical acts. *Journal of Applied Psychology*, *100*(2), 360–375. https://doi.org/10.1037/a0037796

Lau, J. T. F., Yang, X., Tsui, H., & Kim, J. H. (2003). Monitoring community responses to the SARS epidemic in Hong Kong: From day 10 to day 62. *Journal of Epidemiology and Community Health*, *57*(11), 864–870. https://doi.org/10.1136/jech.57.11.864

Lee, S. T., & Basnyat, I. (2013). From Press Release to News: Mapping the Framing of the 2009 H1N1 A Influenza Pandemic. *Health Communication*, *28*(2), 119–132. https://doi.org/10.1080/10410236.2012.658550

Leung, G. M., Ho, L.-M., Chan, S. K. K., Ho, S.-Y., Bacon-Shone, J., Choy, R. Y. L., … Fielding, R. (2005). Longitudinal Assessment of Community Psychobehavioral Responses During and After the 2003 Outbreak of Severe Acute Respiratory Syndrome in Hong Kong. *Clinical Infectious Diseases*, *40*(12), 1713–1720. https://doi.org/10.1086/429923

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*, *19*(5), 469–479. https://doi.org/10.1016/0022-1031(83)90023-9

Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PLoS ONE*, *13*(12), 1–12. https://doi.org/10.1371/journal.pone.0208034

Negd, M., Mallan, K. M., & Lipp, O. V. (2011). The role of anxiety and perspective-taking strategy on affective empathic responses. *Behaviour Research and Therapy*, *49*(12), 852–857. https://doi.org/10.1016/j.brat.2011.09.008

Nerb, J., & Spada, H. (2001). Evaluation of environmental problems: A coherence model of cognition and emotion. *Cognition and Emotion*, *15*(4), 521–551. https://doi.org/10.1080/02699930126254

Nitz, M., & West, H. (2004). Framing of newspaper news stories during a presidential campaign cycle: The case of Bush-Gore in Election 2000. In S. L. Senecah (Ed.), *The environmental communication yearbook* (pp. 99–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Price, V., Tewksbury, D., & Powers, E. (1997). Switching trains of thought: The impact of news frames on readers' cognitive responses. *Communication Research*, *24*(5), 481–506. https://doi.org/10.1177/009365097024005002

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it!

Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology and Behavior*, *130*, 127–134. https://doi.org/10.1016/j.physbeh.2014.04.002

Scheufele, D. A., & Iyengar, S. (2012). The State of Framing Research: A Call for New Directions. *The Oxford Handbook of Political Communication*, (August 2016), 1–27. https://doi.org/10.1093/oxfordhb/9780199793471.013.47

Sheafer, T. (2007). How to Evaluate It: The Role of Story-Evaluative Tone in Agenda Setting and Priming. *Journal of Communication*, *57*(1), 21–39. https://doi.org/10.1111/j.1460-2466.2006.00327.x

Small, D. A., & Loewenstein, G. (2003). Helping a Victim or Helping the Victim: Altruism and Identifiability. *The Journal of Risk and Uncertainty*, *1*, 5–16.

Spielberger, C. D., & Gorsuch, R. L. (1983). *The state trait anxiety inventory for adults: sampler set manual, test, scoring key*. Redwood City Calif: Mind Garden.

Stevens, J. D. (1985). Social Utility of Sensational News: Murder and Divorce in the 1920's. *Journalism Quarterly*, *62*(1), 53.

Tabri, N., Hollingshead, S. J., & Wohl, M. J. A. (2020). *Framing COVID-19 as an Existential Threat Predicts Anxious Arousal and Prejudice towards Chinese People Like*. Preprints. https://doi.org/10.1017/CBO9781107415324.004

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Thomas, B. C., Croft, K. E., & Tranel, D. (2012). Harming kin to save strangers: further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage, *23*(9), 2186–2196. https://doi.org/10.1162/jocn.2010.21591.harming

Trémolière, B., Neys, W. De, & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384. https://doi.org/10.1016/j.cognition.2012.05.011

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.4324/9781912282562

Uribe, R., & Gunter, B. (2007). Are "sensational" news stories more likely to trigger viewers'

emotions than non-sensational news stories?: A content analysis of British TV news. *European Journal of Communication*, *22*(2), 207–228. https://doi.org/10.1177/0267323107076770

Vasterman, P., Yzermans, C. J., & Dirkzwager, A. J. E. (2005). The role of the media and media hypes in the aftermath of disasters. *Epidemiologic Reviews*, *27*, 107–114. https://doi.org/10.1093/epirev/mxi002

Vettehen, P. H., Nuijten, K., & Peeters, A. (2008). Explaining effects of sensationalism on liking of television news stories: The role of emotional arousal. *Communication Research*, *35*(3), 319–338. https://doi.org/10.1177/0093650208315960

Yan, Q., Tang, S., Gabriele, S., & Wu, J. (2016). Media coverage and hospital notifications: Correlation analysis and optimal media impact duration to manage a pandemic. *Journal of Theoretical Biology*, *390*, 1–13. https://doi.org/10.1016/j.jtbi.2015.11.002

Zhao, J., Harris, M., & Vigo, R. (2016). Anxiety and moral judgment: The shared deontological tendency of the behavioral inhibition system and the unique utilitarian tendency of trait anxiety. *Personality and Individual Differences*, *95*, 29–33. https://doi.org/10.1016/j.paid.2016.02.024

Supplementary material

Media articles

Episodic: High harm-content

'A brutal, evil virus'. Family heartbroken as Covid-19 claims life of teenager 24 hours after hospital admittance

On Wednesday afternoon, the family of Joseph Selis said their final goodbyes to him over the phone as the teenager tragically became another victim of Covid-19. Hospitals expect more and more patients suffering with this deadly disease to come through their doors in the next few weeks.

In a statement his father said, 'he had so much ahead of him', 'it was unbearable to not be able to see him when he died. I just couldn't believe what was happening'

Jo was admitted to hospital on Tuesday, after a persistent cough worsened. His family said that by midnight, he was 'unable to breathe and in a lot of pain'. He had no underlying health conditions.

'unable to breathe and in a lot of pain'

His brother Stephan warns, 'this is a brutal, evil virus. It doesn't matter who you are, everyone is at risk'.

His family said they are 'devastated' and 'still trying to process' their tragic loss.

Joseph condition worsened dramatically, after only suffering with a mild temperature on Monday he was struggling to breathe the following day and ventilated within 4 hours. The disease was extremely aggressive, and Joseph's lungs eventually stopped working.

'It was horrible to watch. We weren't able to do anything', Stephan said.

In the early hours of Wednesday morning, Joseph got moved to an isolated ward and the hospital staff told his family they would be unable to stay due to further risks of contracting and spreading the disease. Distraught, Joseph's family returned home unable to do anything further.

The following day, they received the terrible news that Joseph would not make it after suffering respiratory distress for several hours he was not able to fight the virus.

It is becoming clear, that the speed and intensity of this infection is unpredictable, as people of all ages with no underlying health conditions continue to die from the disease.

Sadly, health officials predict that many more lives will be lost as they expect the influx of patients with the disease to rise considerably over the coming weeks.

The message from government health officials remains the same:
Stay home. Protect our health service. Save lives.

Thematic high harm-content

Healthcare staff redistributed to high-risk areas to deal with 'overwhelming surge' of Covid-19 patients

In an unprecedented move from the government, healthcare workers around the country are being re-deployed to population-dense regions, as hospitals struggle to respond to the dramatic influx of coronavirus patients.

Some rural hospitals, however, are also struggling to manage patients that come through their doors, as PPE equipment, ventilators and intensive care specialists are typically much scarcer in these areas.

Dr Penn, a chief medical officer, said hospital capacity has been "pushed to its limits".

In some cases, "up to 45% of healthcare staff have been unable to attend work" due to personal vulnerability or coronavirus symptoms. Healthcare staff face high risks of catching the virus every day, often bringing that risk home to their families.

A central hospital yesterday was forced to announce a 'critical incident' as they struggled to manage new coronavirus patients. A lack of intensive care staff and ventilators meant patients could not receive care soon enough, with many transferred to a nearby hospital.

As the death toll rises, government health officials predict this is only the beginning of the emergency, as an 'overwhelming surge' of patients currently incubating the virus are expected to be admitted in the coming weeks. Heads of emergency response and national Ambulance services have expressed fears they will be unable to respond to the crushing number of patients requiring transport, as emergency callouts are expected to increase by 120 a day.

"Healthcare staff are already exhausted and struggling under the weight of this disease"

The move to spread out healthcare resources, is part of a comprehensive set of measures proposed by senior healthcare officials announced yesterday. The military are expected to continue their efforts in building new hospitals for those suffering with the disease, but also for patients needing life-saving treatment away from the risks of Covid-19.

Jo Selva, a spokesperson for one of the major central hospitals, has said "Healthcare staff are already exhausted and struggling under the weight of this disease. Sadly, the influx of very ill patients coming through their doors in the next few weeks is likely to greatly exceed previous predictions."

The message from government health officials remains the same:

Stay home. Protect our health service. Save lives.

Episodic low harm-content

Facing coronavirus everyday: Growing anxiety of shopkeepers and pharmacists told to go to work

Keyworkers around the country in our supermarkets, pharmacies and drugstores continue to go to work, putting themselves and often their families at risk of catching the virus every day.

Joseph Selis works in a large grocery store and has said he fears for his health constantly.
"It's hard not to worry, when we see the news everyday telling people to stay at home and stay safe. I can't do that."

Although strict social-distancing protocols are in place, Joseph says "it is very difficult to avoid people completely" in supermarkets, as the option of home deliveries is still not available for the majority of people.

Joseph also expressed concern for those around him; "I live with my parents and have a younger brother and sister. I'm scared I'll accidentally spread the virus to them".

This is just the beginning, as many key workers could be asked to continue working for several months while the majority of the country is advised to stay home.

"I can't help but think it is more and more likely I will catch the virus"

Pharmacists and drugstores are also under increasing pressure to remain open. Stephan works in a busy central pharmacy and is concerned about exposure to people who may have the virus over time.

"We need to stay open to ensure people are able to get the medication they need... I know it's very important but as time goes on, I can't help but think it is more and more likely I will catch the virus".

Stephan says he suffered from a bad bout of the flu a few years ago, which meant he was ill for some time. Unfortunately, he is not considered 'high risk' so "has to keep coming to work". He is worried his poor health in the past may mean he is less equipped to fight the virus if he becomes unwell.

These are no doubt worrying times for those working in public-facing roles every day, as growing numbers of bus drivers, taxi drivers and shopkeepers are being admitted to hospital each week with coronavirus.

The message from government health officials remains the same:

Stay home. Protect our health service. Save lives.

Thematic low harm-content

‘Too little, too late’. Testing strategy failed to gather vital data in early days of Covid-19

Following predictions from virology experts warning of the severity of Covid-19, the country has seen its largest spike in coronavirus cases over the past week, and the World Health Organisation (WHO) has condemned the government’s slow approach to testing people for the virus.

A senior spokesperson in the Department for Health made a formal statement this week, acknowledging “more should have been done in the beginning to understand the spread of the virus. We are directing all of our efforts towards getting large numbers of healthcare workers and civilians tested in the coming weeks.”

“We are now playing ‘catch up’ to understand exactly how many people are infected”

Test-building sites around the country have been frantically developing and packaging test kits in preparation for mass-testing, which has been more delayed than they expected. It is likely that demand for tests may exceed the availability of kits, and large amounts of tests will need to be imported from other countries.

Dr Penn, a researcher in one of the leading virology units in the country said “it is a great relief that this is beginning to happen now, but some will say it is too little, too late. We are now playing ‘catch up’ to understand exactly how many people are infected, and how we can effectively manage this virus”.

The WHO has consistently pushed governments to test people very early on, especially in the case of novel diseases like Covid-19, as emerging data provides vital information to inform swift and effective policy change and ensure the most vulnerable people are protected.

Many countries have been slow to test people for Covid-19, in part due to mathematical modelling research using early projection data that was unable to account for the sheer number of people who would develop severe cases of the disease.

Jo Selva, a senior public health official, stated earlier today “we have been making very difficult decisions in unknown territory which we must learn from. Policy is now rapidly shifting towards a testing focus, to better understand the scale and spread of infection”.

The message from government health officials remains the same:

Stay home. Protect our health service. Save lives.

Moral dilemmas

1.You are a physician in a busy hospital where staff and resources are scarce and need to be rationed. On your ward, there is a patient using a ventilator who is unable to breathe properly without it. Two other patients have recently joined the ward, and are also struggling to breathe, but there are no more ventilators available. Without a ventilator soon, the condition of these two new patients will worsen and they could both be left unable to breathe while waiting. These two patients could manage their coughing and breathlessness by sharing the one ventilator of the first patient.

Do you take the ventilator away from the one patient to share between the other two patients?

2.You have escorted an elderly relative to Accident and Emergency at your local hospital at 6 am in the morning with breathing difficulties and a high temperature and are asked to sit in the waiting room as there are no staff available. At 7:30 am two further elderly patients arrive to A and E who are experiencing identical symptoms of the same severity. A doctor arrives and says the fairest way to choose who is seen next is by random selection.

Do you insist to be seen first?

3.There are 5 very ill coronavirus patients in a waiting room, all requiring urgent attention. One of the patients however has further underlying health difficulties which increases their health risks without treatment. You are the only doctor available in the emergency room.

Do you treat the patient with underlying health difficulties first?

4.You take one of your family members to hospital to receive treatment for a prolonged fever and persistent coughing that is getting worse. A lead nurse in the hospital is also suffering with a fever and persistent coughing and is currently unable to do her job until she receives treatment. There are long wait times and limited staff available.

Do you insist the nurse receives treatment first?

5.There is an elderly patient and a small child who have both come into hospital with a fever and complications from contracting coronavirus that means they require immediate care from doctors and nurses. The small child requires care from their parents and is unable to self-isolate completely during their stay in hospital which could mean the virus is spread to others.

Staff numbers are low in the department, and as the lead medical professional you need to decide who receives care first which could mean that the other patients condition worsens while they wait.

Do you treat the child first and risk the condition of the elderly patient worsening?

6. You have escorted a close friend who lives in a very rural area to a city hospital to receive treatment for severe flu symptoms and cough. They have no other relatives/friends nearby and no other means of transport, and you suspect they may have coronavirus. Your friend is ventilated within a few hours to support their breathing. After a few hours, the doctor asks whether they could remove the ventilator from your friend for a short time, so that another patient may share this resource.

Do you insist the ventilator remains with your friend?

Data reduction and additional analyses

Data reduction

Partially completed questionnaires were removed from the analysis. Statistical analysis was carried out with SPSS v.26. A Pearson's bivariate correlation analysis including all of the key independent and dependent variables was conducted. The SPSS scripts for moderation, mediation and conditional process analyses (PROCESS) were adopted from Hayes (2018). For all moderation analyses carried out in PROCESS, interactions are probed at the 16th, 50th and 84th percentiles by default. The Johnson-Neyman technique implemented in PROCESS was used to identify regions of significance for moderation effects as recommended by Tabachnick and Fidell (2013). A bootstrapping method was adopted for all PROCESS regression analyses (5000 x samples, 95% confidence interval) (Hayes, 2018a). Leverage points and influential data points with standardised residuals ± 3 standard deviations were investigated. No outliers were removed from the analysis. Homogeneity of variances was determined by Levene's test for equality of variances. Homoscedasticity was assessed by visual inspection of standardised residuals plotted against the predicted values. In H2, the assumption of normality for an Analysis of Variance (ANOVA) as assessed by a Shapiro-Wilk test was violated for the dependent variable 'anxiety' in the high harm-content thematic group and for the dependent variable 'subjective harm salience' in the episodic low harm-content group. Moral appropriateness judgments in R2 also demonstrated non-normal distribution and Levene's test of equality of variances was violated. As groups were close in size and ANOVA's considered relatively robust to deviations of normality the dependent variables were not transformed.

Scale reliabilities

Internal reliabilities for interoception subscale 'noticing' (Cronbach's $\alpha = .792$), 'not distracting' ($\alpha = .831$), and 'not worrying' ($\alpha = .753$) all indicated good internal consistency. Similarly, the state anxiety scale indicated very good internal reliability ($\alpha = .950$).

Correlations

Anxiety

Sex (r=.223, p=.006) and age (r=-.268, p=.001) showed significant correlation with anxiety ratings, with older people typically reporting lower levels of anxiety and women reporting higher levels of anxiety than men. Greater anxiety after reading the articles was associated with higher ratings of harm-salience for the news articles (r=.258, p=.001). Therefore, people who found the harm and suffering of others in the news stories to be more striking or salient, also experienced greater levels of anxiety.

Subjective harm-salience

SHS did not show any correlation with age (r=-.016, p=.851) or sex (r=-.021, p=.802), but was positively associated with subjective deadliness of COVID-19 (r=.161, p=.048). As this measure correlated with the interoception (not worrying) subscale, it was not included in the ANCOVA.

Interoception

The interoception subscales ('noticing', 'not distracting' and 'not worrying') did not show any correlation with each other, confirming the independence of these constructs. People scoring lower on the 'not worrying' subscale were more likely to report high levels of anxiety (r=-.291, p<.005); perceive coronavirus to be more threatening (r=-.224, p=.006); and believed it to be more likely that they or someone they knew would die from the disease (r=-.195, p=.016). This is in line with previous findings (Mehling et al., 2018) and suggests that people who typically worried about unpleasant or painful bodily sensations had more extreme health concerns surrounding the virus and experienced greater anxiety after reading the COVID-19 news articles. The 'noticing' and 'not distracting' subscales were not correlated with anxiety. People scoring lower on the 'not distracting' subscale were more likely to report that they would find it unpleasant staying at home during lockdown (r=.268, p=.001), suggesting people who were less able to be present with bodily sensations expected it to be more unpleasant to be confined to their own homes.

Individual and demographic factors associated with behavioural intentions

We explored individual and demographic factors associated with preventative, prosocial and antisocial behaviours with a Pearson's bivariate correlation. Women were significantly more likely to carry out preventative (r=.322, p<.0005) and prosocial behaviours (r=.173, p=.033), with men more likely to carry out antisocial behaviours (r=-.164, p=.045). People who perceived a greater threat from COVID-19 were more likely to carry out preventative behaviours (r=.347, p<.0005). People finding it more unpleasant to stay at home were less likely to carry out preventative behaviours (r=-.207, p=.011) and more likely to carry out antisocial behaviours such as breaking protocol or bulk buying (r=.221, p=.007). People who knew someone who had tested positive with the virus were less likely to carry out antisocial behaviours (r=-161, p=.049). Age, number of cases nearby, and subjective deadliness of COVID-19 did not predict any of the behavioural measures. Overall, these findings suggest that women were more likely to carry out behaviours to keep themselves and others safe during the pandemic. Unpleasant experiences of staying at home are also likely to contribute to behaviours that could heighten people's risk of catching and spreading the virus.

Principal Component analyses

For both of the Principal Component analyses, Bartletts test of sphericity was significant (p<.0005), suggesting the data was appropriate for factor analysis. The Kaiser-Meyer-Olkin (KMO) measure was .639, a mediocre level of sampling adequacy (Kaiser & Rice, 1974).

Mediating role of SHS in relationship between interoception and anxiety

A mediation analysis using PROCESS (Model 4, Hayes, 2018) was implemented to investigate whether interoception indirectly affected anxiety by influencing appraisals of harm salience. We found interoception directly influenced anxiety ratings (t=-3.380, p=<.0005, 95% CI: -6.064 to -1.589) and was not mediated by SHS, as confidence intervals for the indirect effect included zero (-1.0319 to .0448). Moreover, a simple moderation analysis (Model 1: Hayes, 2018), showed the effect of SHS on anxiety ratings was not moderated by interoception (not worrying) (R^2 *chng* =.0001, F (1, 147) =.0141, p=.9058). This suggests that a tendency to worry about bodily sensations and subjective appraisals of harm-salience independently increased anxiety ratings.

Study 4 Postface

Paper:

Brown, H., Fraser, D. S., Farmer, H., & Proulx, M. J. (2020). *The salience of harm: the framing and harm-content of coronavirus media articles indirectly influences preventative and antisocial behaviours during the pandemic [Unpublished thesis].* University of Bath.

The key finding from this study was that subjective appraisals of harm salience influenced behavioural intentions associated with the pandemic. Harm salience has not been captured in previous studies exploring media-effects and suggests a promising avenue for future public health research during pandemics. This measure conceptualised how salient or striking people found the harm and suffering of those in the article to be and indicates that our emotional response to the suffering of others during this pandemic could be strong enough to change behaviours that impact public health. In particular, we found that highly harm-salient thematic narratives were just as impactful for appraisals of harm-salience as episodic narratives. This suggests that narratives need not pull at the heart strings of readers with narrowly focused, evocative stories of individuals to generate powerful emotional responses from readers. Thematic narratives that aim to provide useful contextual information regarding the coronavirus crisis in wider society, may be as effective for emphasising harm on a larger scale whilst also providing audiences with the 'bigger picture'.

The finding that the 'not worrying' subscale of interoception influenced subjective appraisals of harm salience, highlighted a potentially important individual differences factor in how people respond to health-crisis information in the media. People who typically worry about unpleasant or painful bodily sensations, reported greater anxiety after reading the article and found the harm associated with coronavirus to be more salient when controlling for the frame-type and harm-content of the articles. This suggests that how much people worry about their own bodily sensations has implications for how they interpret the harm and suffering of others, but it is unclear whether this is an empathic response to the suffering of others or a self-oriented distress that is triggered by highly salient illness-related media. As subjective harm salience and anxiety emerged as key factors influencing preventative and antisocial behavioural intentions during the pandemic, the role of interoceptive sensibility in health behaviours warrants much further investigation. Anxiety appears to provide a useful

self-oriented emotional response encouraging people to protect themselves, and indirectly protecting others from viral spread during the pandemic. However, anxiety did not appear to provide any benefit for prosocial behaviours, or for the moral dilemma task. This is consistent with Studies 1-3, that did not find anxiety to be associated with moral judgments or behaviour that directly concern the wellbeing of others. It is possible that there are more complex interactions between anxiety, interoception and moral judgment that may come to light in study designs using clinical populations or those involving direct manipulations of anxiety.

The findings from this study suggest the saliency of harm can be enhanced in text-based media and may have implications for real-world behaviours during the pandemic. However, harm salience was not associated with moral judgments about the treatment of coronavirus patients, suggesting that appraisals of harm and suffering of others may not influence moral judgments that are hypothetical and thus inconsequential in regard to the pandemic. Further work is needed to establish the mechanism of appraisals of harm-salience in more real-world health contexts, and in moral dilemma research to understand whether people's subjective appraisals of harm are associated with traits such as action aversion and outcome aversion (Miller et al., 2014) and response tendencies of harm-aversion and outcome-maximisation (Conway & Gawronski, 2013).

Conclusion

Damasio (1996) proposed that somatic signals and homeostatic states in the body are intrinsically linked to social decision-making because personal and social issues are strongly tied to outcomes of punishment and reward. 'Good' and 'bad' are felt physically and emotionally as pleasure and pain which are represented, alongside homeostatic states within the somatosensory system (Damasio, 1996). We have found that individual differences in interoception can interact with emotional and physiological processes to influence moral judgments and behaviour in numerous ways. In particular interoceptive accuracy, moderated the relationship between physiological arousal and moral judgments of harm. Interoceptive accuracy also predicted a lengthening of pre-ejection period in VR moral dilemmas, which was also associated with differences in response time to carry out harmful action. Interoceptive sensibility indirectly influenced harm-aversion responses to a moral dilemma task, and a tendency to worry about bodily sensations increased appraisals of harm salience of coronavirus media articles, which predicted preventative and antisocial behavioural intentions during the pandemic. These findings provide further support for the Somatic Marker Hypothesis (Damasio, 1996) and are consistent with evidence from studies with patients with VMPfC damage, that show how the physiological representation of emotional states may influence harm-based moral decision-making (Damasio et al.,1990; Koenigs et al., 2007; Moretto et al., 2010; Young et al., 2010). The lack of association found between distinct measures of interoception in study 2 and 3, also provide further support for the dimensional model of interoception developed by Garfinkel and Critchley (2013).

This thesis builds on previous work, by establishing how individual differences in interoceptive capacities in healthy populations can influence harm-based moral decision-making, indicating that the relationship between bodily signals and moral decision-making may be stronger for some people than others. Perhaps most significantly, our findings suggest that an ability to consciously direct attention to heartbeats can modulate the relationship between changes in physiological arousal and moral judgments and behaviour. Furthermore, this work adds to previous research using traditional moral dilemma paradigms (Greene et al., 2001, 2004; Sarlo et al., 2014), by using moral dilemma stimuli that can establish harm-aversion and outcome-maximisation response tendencies or inclinations (Conway & Gawronski, 2013) as well as contrasting an allocentric measure of moral judgment. This allowed a more nuanced investigation of how emotional and interoceptive processes may be

linked with distinct socio-emotional motivations driving choice judgments (i.e. Would you…?), and acceptability judgments of harmful acts (i.e. How acceptable do you find…?) (Tassy et al., 2013). We have also shown for the first time, how a sub-construct of interoceptive sensibility may present a dispositional vulnerability to harm-salient news media about coronavirus, which may influence social behaviour during the pandemic with consequences for viral spread. Study 3 and 4 compliment findings from study 1 and 2 using text-based moral dilemma paradigms, to show how interoceptive processes may be associated with moral behaviour in more ecologically valid contexts. Although the cross-sectional and behavioural study designs used, prevent us from drawing conclusions about the neural bases of the effects found, this work can complement neuropsychological research linking the processing of somatic markers in the brain to observable phenomena using moral dilemma paradigms, such as those measuring event-related potentials while carrying out moral dilemma tasks (Sarlo et al., 2014; Yoder & Decety, 2014).

Summary of findings

In Study 1, we found that a tendency to notice internal sensations (interoceptive sensibility) indirectly increased harm-aversion response tendencies on a moral dilemma task, which was mediated by a greater tendency to provide 'intuitive' answers when faced with cognitively demanding problems. The complexity and length of the moral dilemma stimuli used potentially combined with the low task-engagement of people completing online research, may be partly responsible for this effect. People with a greater tendency to rely on intuitive heuristics, may be more likely to support harmful action on congruent (i.e. non-moral conflict dilemmas) if they have not fully considered the context and outcomes of the problem. In addition, the interoceptive state of hunger appeared to be uniquely influential for allocentric judgments of non-utilitarian harmful acts, suggesting that self-reports of hunger-type sensations may predict greater acceptance of harmful acts when in the role of observer, potentially due to an absence of disgust sensations that may be triggered by the release of satiety hormones (Halawi et al., 2017; Vicario et al., 2018).

In Study 2, we found several moderation effects of interoceptive accuracy on the relationship between physiological arousal and moral judgments. Notably, on incongruent trials, a greater increase in cardiovascular indices of sympathetic arousal (threat reactivity) predicted reduced harm-aversion tendencies, whereas, increased heart rate predicted greater harm-aversion tendencies and decreased heart rate predicted outcome-maximisation

tendencies, but only for people who were relatively better at perceiving heartbeats. Although we are not able to establish the causal mechanisms discussed in predictive coding models of interoception (Apps & Tsakiris, 2014; Ainley et al., 2016; Farb et al., 2015; Seth & Friston, 2016) that may underlie these effects, we speculate that heartbeat detection ability may selectively modulate the influence of changes in physiological arousal on moral decision-making. Possibly due to a greater ability to enhance the precision of cardiovascular signals with attention (Ainley et al., 2016), better heartbeat perceivers may demonstrate greater non-reactivity to surprising interoceptive sensations afforded by perceptual forms of inference that prioritise precision of bodily information (Farb et al., 2015). A greater sensitivity to interoceptive sensations may facilitate adaptive regulation strategies (Füstös et al., 2013; Kever et al., 2015) that temper overt behavioural responses to changes in physiological arousal; selectively attenuating or strengthening the relationship between physiological arousal and moral judgments.

In Study 3, a virtual-reality moral dilemma study, we found interoceptive accuracy was associated with relatively reduced sympathetic cardiovascular arousal (pre-ejection period; PEP) prior to taking harmful action on incongruent trials. PEP reactivity was associated with different response time patterns, depending on the nature of the harmful action taken in a semi-autonomic vehicle (button-press versus foot-pedal). This suggests that a superior ability to detect heartbeats could influence response time to carry out harmful action in moral dilemma driving tasks. Further work is needed to understand whether reduced PEP reactivity demonstrated by better heartbeat detectors is due to a relatively greater motivational challenge state (Tomaka et al., 1993), or reduced physical and cardiovascular effort to carry out the harmful actions (Herbert et al., 2007). If the latter is true, capturing individual differences in interoception would be pertinent for future psychophysiological studies exploring moral behaviour in VR environments. Study 2 and 3 contribute to literature exploring the role of embodiment and personal harm in VR moral dilemmas (Francis et al., 2016; Francis et al., 2017) and psychophysiological studies exploring the role of sympathetic cardiovascular arousal in action-based aversion to harm (Cushman et al., 2012; Parton & McGinley, 2019), by showing that changes in physiological arousal generated by the prospect of harming others, may be moderated by an ability to perceive cardiac sensations.

Although we found some interesting significant relationships between interoception and moral decision-making in studies 1-3, we did not find evidence that any of interoception

measures directly influenced moral decision-making or behaviour. The effects found were either moderated or mediated, suggesting that individual differences in interoception alone are unlikely to be useful to predict moral decision-making in moral dilemmas of harm. In addition, our measures of interoceptive sensibility and interoceptive meta-cognitive awareness did not influence moral decision-making or moderate the influence of emotional arousal on decision-making in studies 1-3. Gastric sensitivity using the water load task, also did not show any relation to the moral behaviour outcomes in study 3 and did not interact with any of the primary dependent variables in study 2. The absence of relationships found between these interoception parameters indicates that the relationship between physiological arousal and moral decision-making is uniquely influenced by individual differences in interoceptive accuracy. We can take from this that a tendency to notice bodily sensations and an awareness of one's ability to detect heartbeat sensations may be less relevant to moral decision-making processes when considering moral judgments of harm. However, further work is needed to understand whether dimensions of interoception, potentially exploring and comparing other measures of interoception, could interact with each other to directly influence moral decision-making alone or modulate the link between emotional arousal and moral decision-making.

Finally, in study 4, we found that subjective appraisals of harm salience of coronavirus media content and feelings of anxiety predicted preventative, prosocial and antisocial behavioural intentions during the pandemic. Greater anxiety and harm-salience were associated with a tendency to worry about painful or unpleasant bodily sensations - a sub-construct of interoceptive sensibility (Mehling et al., 2012). This suggests that when exposed to harm-salient media coverage, a negative attentional focus on bodily sensations could shape how people feel and behave, which could have implications for viral spread in useful ways (preventative behaviours) and in less useful ways (antisocial behaviours). Media-coverage of the coronavirus pandemic is inherently political, and these findings are pertinent to consider in the context of 'visceral politics' (Tsakiris, 2020). How harm is made salient in the media, potentially as a tool to emphasise normative guidance around how we 'should' be conducting ourselves during the pandemic, has implications on a societal scale for how safe people feel and the behaviours they are likely to engage in (Tsakiris, 2020).

Implications and future research

There is much scope to explore the role of interoception across other types of moral decision-making and behaviour in different contexts. The role of disgust sensations and disgust sensitivity has already been highlighted as a factor influencing judgments of others ethical violations of harm which may be triggered by homeostatic changes associated with satiety (Vicario et al., 2018), which may also be modulated by interoceptive sensibility (Schnall et al., 2008). Feelings of disgust have shown to be associated with activation of the insula cortex (Wright et al., 2004), which is a key area implicated in interoceptive and emotional experiences (Zaki et al., 2012). Future research using other measures of interoception, potentially alongside measures of gastric myoelectrical activity could provide insight into the mechanisms underlying these effects. It may also shed light on why cardiac and gastric forms of interoception were associated with moral judgments in different ways, and not associated with each other in the current study and also in a recent study that did not find support for a generalised interoceptive ability across channels (Ferentzi et al., 2018a). In addition, other research has found different effects of emotional regulation difficulties, arousal and valence on moral judgments across the domains of Harm, Fairness, Authority, Loyalty and Sanctity (Zhang et al., 2017). Therefore, individual differences in emotional regulation and perhaps interoception, may influence moral judgments in other moral domains, depending on the nature and strength of the emotion experienced.

Secondly, although interoceptive accuracy has shown to be a relatively stable trait (Ferentzi et al., 2018b), there is evidence to suggest that interoceptive accuracy may be momentarily enhanced, for example using direct gaze to heighten self-awareness (Isomura & Watanabe, 2020) or use of biofeedback (Meyerholz et al., 2019). Therefore, interoceptive accuracy could present a promising 'leverage point' to alter the relationship between somatic states and moral decision-making in real-world contexts. In particular, heartbeat detection ability using a heartbeat counting task has been associated with antisocial behaviour in forensic populations (Nentjes et al., 2013). Although individual, genetic and environmental factors contributing to criminal behaviour are undoubtedly complex and interrelated, it is possible that enhancing people's ability to perceive somatic sensations associated with punishment and reward (Damasio, 1996), if sustained in the longer term, could have positive behavioural outcomes among forensic populations. Importantly, Nentjes et al (2013) highlight that the cross-sectional nature of the study means it is not possible to determine

causality in the relationship between expressions of antisocial behaviour and heartbeat detection ability. Developmental studies exploring interoception and moral decision-making and behaviour in children may shed light on how interoceptive processes are related to psychopathology and anti-social behaviour over time. A recent study found overweight adolescents showed atypical insular activation prior to risky decision making, specifically, decreased activation of brain regions associated with risk, and increased activation of brain regions associated with reward during a Risky-Gains task (Delgado-Rico et al., 2013). This shows how the processing of visceral signals may influence complex decision-making at a young age. Heartbeat detection ability in adults has also been associated with enhanced decision-making ability in the Iowa Gambling task (Werner et al., 2009). Further work is needed to understand how interoceptive processes in childhood may be associated with moral decision-making, and whether interoceptive capacities of children such as heartbeat detection ability, may be more malleable or receptive to interventions that improve sensitivity to somatic sensations.

Thirdly, investigating the role of interoceptive processes and experiences of presence in conjunction with emotional and physiological processes in moral dilemma tasks, may further shed light on the distinct emotional and psychological processes governing moral judgment and action that were discussed by Francis et al (2016, 2017) in their comparisons of text-based and virtual reality moral choices. Understanding the difference in homeostatic drivers and interoceptive processes recruited during text-based versus immersive moral dilemma scenarios, may also be useful to illuminate the regulatory mechanisms that underlie physiological and behavioural responses to harm-based moral dilemmas. Further investigations could also clarify whether performance on heartbeat discrimination tasks (Whitehead et al., 1977) show the same relationships as heartbeat detection ability with moral judgments and arousal found in Study 2 and 3. In particular, using this task may provide more understanding of whether a person's ability to integrate cardiac sensory information with external sensory information (Forkmann et al., 2016) is important in the relationship between physiological arousal and moral decision-making.

Furthermore, VR moral dilemma studies investigating individual differences in physiological responses preceding emergency situations, offers a promising methodology for testing artificial intelligence technology currently being trialled in autonomous vehicles, that incorporate driver data into crash prediction algorithms. Ethically, it is not possible to

comprehensively explore the behaviour of drivers in real collision events on the road. VR simulations and advanced car simulators may be the best we can do to better understand factors that predict harm-avoidance or 'anti-utilitarian' behaviours of drivers in automated vehicles. It also allows the opportunity to investigate the impact of novel automated functions and driving actions (such as swopping an accelerator pedal with a button) in automated vehicles, in more ecologically valid and sensorially rich environments that offer a greater level of experimental control than field experiments. Future research measuring real-time physiological data and moral behaviour of drivers in VR driving-dilemmas, could provide useful sample data to inform supervised-learning models of AI (Ba et al., 2017) and physiological classifier systems in SA vehicles (Veeraraghavan et al., 2007) that are designed to calculate crash probabilities in real-time.

Finally, future research could explore interactions between interoception, presence and agency in VR moral dilemmas. Agency is defined as the experience of being in control of one's body and the external world. Predictive coding models have linked experiences of agency, interoceptive processing and conscious experiences of presence (Seth, 2013; Seth et al., 2012; Farb et al, 2015), which could be informative for understanding individual differences in moral behaviours carried out in VR moral dilemmas. Issues of agency are also particularly relevant to consider in the context of human-computer interactions as the design of these interactions can influence both explicit and implicit forms of agency (see Limerick et al., 2014). When considering the design of human-computer interactions, such as the interfaces used in autonomous vehicles, it would be worthwhile to investigate whether the nature of an interaction that could have consequences for the wellbeing of others (such as novel driving actions), interferes with experiences of agency and also a sense of moral responsibility of human supervisors who ultimately bare legal accountability for these vehicles (Limerick et al., 2014).

In conclusion, this thesis has demonstrated that individual differences in interoception can influence responses to harm-based moral decision-making, suggesting that the emotional processes believed to precipitate moral judgments of harm (e.g. Cushman et al., 2012; Greene et al., 2001; Reynolds & Conway, 2018) may depend on how we perceive and interpret somatic signals in our body (Damasio, 1996). People who are better at perceiving their heartbeats may generate heightened physiological responses to aversive moral dilemma stimuli. Interoceptive accuracy also appears to moderate the relationship between

cardiovascular indices of arousal and harm-aversion, outcome-maximisation in text-based paradigms and response time for harmful action in VR dilemmas. In addition, individual differences in how much people worry about painful or unpleasant sensations within their body may influence how they respond to harm-salient media about coronavirus, and the socially relevant behaviours they intend to carry out during the pandemic and suggests this may be a worthwhile construct to explore in the context of socially-relevant health behaviours. Further work is needed to establish how interoceptive processes may interact with emotional and physiological processes to influence moral judgment and behaviour across a range of domains and contexts. Future VR moral dilemma studies incorporating measures of interoception have a lot to offer in terms of understanding how interoceptive processes may interact with arousal, and reports of conscious presence and agency (Seth et al., 2012) to influence how people respond in more 'real-life' moral dilemma scenarios.

References

Adolphs, R., Russell, J. A., & Tranel, D. (1999). A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. *Psychological Science*, *10*(2), 167–171. https://doi.org/10.1111/1467-9280.00126

Ainley, V., Apps, M. A. J., Fotopoulou, A., & Tsakiris, M. (2016). "Bodily precision": A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0003

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia*, *60*(1), 21–28. https://doi.org/10.1016/j.neuropsychologia.2014.05.010

Ainley, V., Maister, L., Brokfeld, J., Farmer, H., & Tsakiris, M. (2013). More of myself: Manipulating interoceptive awareness by heightened attention to bodily and narrative aspects of the self. *Consciousness and Cognition*, *22*(4), 1231–1238. https://doi.org/10.1016/j.concog.2013.08.004

Ainley, V., & Tsakiris, M. (2013). Body Conscious? Interoceptive Awareness, Measured by Heartbeat Perception, Is Negatively Correlated with Self-Objectification. *PLoS ONE*,

8(2). https://doi.org/10.1371/journal.pone.0055568

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Ba, Y., Zhang, W., Wang, Q., Zhou, R., & Ren, C. (2017). Crash prediction with behavioral and physiological features for advanced vehicle collision avoidance system. *Transportation Research Part C: Emerging Technologies*, *74*, 22–33. https://doi.org/10.1016/j.trc.2016.11.009

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801. https://doi.org/10.1037/xge0000533

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, *31*(2), 101–119. https://doi.org/10.1080/0305724022014322

Barlow, D. H. (1988). *Anxiety and its disorders*. New York: Guilford Press.

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*(3), 525–537. https://doi.org/10.1037/0022-3514.77.3.525

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372. https://doi.org/10.1016/j.geb.2004.06.010

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, *275*(5304), 1293 LP – 1295. http://science.sciencemag.org/content/275/5304/1293.abstract

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, *6*(2), 215–225. https://doi.org/10.1093/cercor/6.2.215

Behnke, M., & Kaczmarek, L. D. (2018). Successful performance and cardiovascular markers of challenge and threat: A meta-analysis. *International Journal of Psychophysiology*, *130*(February), 73–79. https://doi.org/10.1016/j.ijpsycho.2018.04.007

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage*, *31*(2), 945–950. https://doi.org/10.1016/j.neuroimage.2005.12.039

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *J.P. Forgas (Ed.) Feeling and thinking: The role of affect in social cognition* (Ed., pp. 59–82). Cambridge University Press.

Boeckxstaens, G. E., D.P., H., Berkhout, B., & Tytgat, G. N. (1990). Is a drink test a valuable tool to study proximal study function? *Gastroenterology*, *116*(A960).

Bornemann, B., Herbert, B. M., Mehling, W. E., & Singer, T. (2015). Differential changes in self-reported aspects of interoceptive awareness through 3 months of contemplative training. *Frontiers in Psychology*, *5*, 1504. https://doi.org/10.3389/fpsyg.2014.01504

Botvinick, M. M., & Rosen, Z. B. (2009). Anticipation of cognitive demand during decision-making. *Psychological Research*, *73*(6), 835–842. https://doi.org/10.1007/s00426-008-0197-8

Braithwaite, J. J., & Watson, D. G. (2015). Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs). *Technical Research Note.* Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham ©

Brener, J., Liu, X., & Ring, C. (1993). A method of constant stimuli for examining heartbeat detection: Comparison with the Brener-Kluvitse and Whitehead methods. *Psychophysiology*, *30*(6), 657–665. https://doi.org/10.1111/j.1469-8986.1993.tb02091.x

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0015

Bults, M., Beaujean, D. J. M. A., De Zwart, O., Kok, G., Van Empelen, P., Van Steenbergen,

J. E., Richardus, J. H., & Voeten, H. A. C. M. (2011). Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: Results of three consecutive online surveys. *BMC Public Health*, *11*, 1–13. https://doi.org/10.1186/1471-2458-11-2

Buon, M., Seara-Cardoso, A., & Viding, E. (2016). Why (and how) should we study the interplay between emotional arousal, Theory of Mind, and inhibitory control to understand moral cognition? *Psychonomic Bulletin and Review*, *23*(6), 1660–1680. https://doi.org/10.3758/s13423-016-1042-5

Carmona-Perera, M., Reyes del Paso, G. A., Pérez-García, M., & Verdejo-García, A. (2013). Heart rate correlates of utilitarian moral decision-making in alcoholism. *Drug and Alcohol Dependence*, *133*(2), 413–419. https://doi.org/10.1016/j.drugalcdep.2013.06.023

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology*, *43*(3), 489–496. https://doi.org/10.1016/j.jesp.2006.05.003

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1249–1264. https://doi.org/10.1016/j.neubiorev.2012.02.008

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235. https://doi.org/10.1037/a0031021

Corley, M. C. (2002). Nurse moral distress: A proposed theory and research agenda. *Nursing Ethics*, *9*(6), 636–650. https://doi.org/10.1191/0969733002ne557oa

Craig, B. A. D. (2015). *How do you Feel?:An Interoceptive Moment with Your Neurobiological Self*. Princeton University Press.

Critchley, H. D., & Garfinkel, S. N. (2018). The influence of physiological signals on cognition. *Current Opinion in Behavioral Sciences*, *19*, 13–18. https://doi.org/10.1016/J.COBEHA.2017.08.014

Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, *77*(4), 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*(2), 189–195. https://doi.org/10.1038/nn1176

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: the aversion to harmful action. *Emotion (Washington, D.C.)*, *12*(1), 2–7. https://doi.org/10.1037/a0025071

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Damasio, A. R. (1994). *Descartes' error: Emotion, Reason and the Human Brain.* Grosset/Putnam books.

Damasio, A. R. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

De Schepper, H. U., Cremonini, F., Chitkara, D., & Camilleri, M. (2004). Assessment of gastric accommodation: overview and evaluation of current methods. *Neurogastroenterology*, *16*(3), 275–285.

Delgado-Rico, E., Soriano-Mas, C., Verdejo-Román, J., Río-Valle, J. S., & Verdejo-García, A. (2013). Decreased insular and increased midbrain activations during decision-making under risk in adolescents with excess weight. *Obesity*, *21*(8), 1662–1668. https://doi.org/10.1002/oby.20375

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior*, *32*(3), 204–215. https://doi.org/10.1016/j.evolhumbehav.2011.01.003

Di Lernia, D., Serino, S., Pezzulo, G., Pedroli, E., Cipresso, P., & Riva, G. (2018). Feel the time. Time perception as a function of interoceptive processing. *Frontiers in Human Neuroscience*, *12*(March), 1–17. https://doi.org/10.3389/fnhum.2018.00074

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R.,

Lawrence, A. D., & Dalgleish, T. (2010). Listening to Your Heart: How Interoception Shapes Emotion Experience and Intuitive Decision Making. *Psychological Science*, *21*(12), 1835–1844. https://doi.org/10.1177/0956797610389191

Duschek, S., Werner, N. S., Reyes del Paso, G. A., & Schandry, R. (2015). The Contributions of Interoceptive Awareness to Cognitive and Affective Facets of Body Experience. *Journal of Individual Differences*, *36*(2), 110–118. https://doi.org/10.1027/1614-0001/a000165

Ehlers, A., Breuer, P., Dohn, D., & Fiegenbaum, W. (1995). Heartbeat perception and panic disorder: possible explanations for discrepant findings. *Behaviour Research and Therapy*, *33*(1), 69–76. https://doi.org/10.1016/0005-7967(94)E0002-Z

Eichler, S., Katkin, Edward, S., Blascovich, J., & Kelsey, R. M. (1987). Cardiodynamic factors in heartbeat detection and the experience of emotion [Abstract]. *Psychophysiology*, *24*, 587.

Eichler, S., & Katkin, E. S. (1994). The relationship between cardiovascular reactivity and heartbeat detection. *Psychophysiology*, *31*(3), 229–234. https://doi.org/10.1111/j.1469-8986.1994.tb02211.x

Fairclough, S. H., & Goodwin, L. (2007). The effect of psychological stress and relaxation on interoceptive accuracy: Implications for symptom perception. *Journal of Psychosomatic Research*, *62*(3), 289–295. https://doi.org/10.1016/j.jpsychores.2006.10.017

Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., & Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in Psychology*, *6*(JUN), 1–26. https://doi.org/10.3389/fpsyg.2015.00763

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2018). Human Decisions in Moral Dilemmas are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-018-0020-x

Feldman Barrett, L. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Feldman Barrett, L., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional. *J Pers Soc Psychol.*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Ferentzi, E., Bogdány, T., Szabolcs, Z., Csala, B., Horváth, Á., & Köteles, F. (2018a). Multichannel investigation of interoception: Sensitivity is not a generalizable feature. *Frontiers in Human Neuroscience*, *12*(June), 1–9. https://doi.org/10.3389/fnhum.2018.00223

Ferentzi, E., Drew, R., Tihanyi, B. T., & Köteles, F. (2018b). Interoceptive accuracy and body awareness – Temporal and longitudinal associations in a non-clinical sample. *Physiology and Behavior*, *184*(August 2017), 100–107. https://doi.org/10.1016/j.physbeh.2017.11.015

Ferentzi, E., Köteles, F., Csala, B., Drew, R., Tihanyi, B. T., Pulay-Kottlár, G., & Doering, B. K. (2017). What makes sense in our body? Personality and sensory correlates of body awareness and somatosensory amplification. *Personality and Individual Differences*, *104*, 75–81. https://doi.org/10.1016/j.paid.2016.07.034

Fissler, M., Winnebeck, E., Schroeter, T., Gummersbach, M., Huntenburg, J. M., Gaertner, M., & Barnhofer, T. (2016). An Investigation of the Effects of Brief Mindfulness Training on Self-Reported Interoceptive Awareness, the Ability to Decenter, and Their Role in the Reduction of Depressive Symptoms. *Mindfulness*, *7*(5), 1170–1181. https://doi.org/10.1007/s12671-016-0559-z

Foot, P. (2003). The Problem of Abortion and the Doctrine of the Double Effect. *Virtues and Vices*, *5*, 19–32. https://doi.org/10.1093/0199252866.003.0002

Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., & Schulz, A. (2016). Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology*, *109*(November), 71–80. https://doi.org/10.1016/j.ijpsycho.2016.09.019

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2018). Virtual morality in the helping professions: Simulated action and resilience. *British Journal of Psychology*, 109(3), 442-465

Francis, K.B., Howard, C., Howard, I. S., Gummerum, M., Anderson, G. G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS*

ONE, *11*(10). https://doi.org/10.1371/journal.pone.0164374

Francis, K.B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating Moral Actions: An Investigation of Personal Force in Virtual Moral Dilemmas. *Scientific Reports*, *7*(1), 13954. https://doi.org/10.1038/s41598-017-13909-9

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. https://doi.org/10.1257/089533005775196732

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Furedy, J. J., & Gagnon, Y. (1969). Relationships between and sensitivities of the galvanic skin reflex and two indices of peripheral vasoconstriction in man. *Journal of Neurology, Neurosurgery, and Psychiatry*, *32*(3), 197–201. https://doi.org/10.1136/jnnp.32.3.197

Füstös, J., Gramann, K., Herbert, B. M., & Pollatos, O. (2013). On the embodiment of emotion regulation: Interoceptive awareness facilitates reappraisal. *Social Cognitive and Affective Neuroscience*, *8*(8), 911–917. https://doi.org/10.1093/scan/nss089

Gailliot, M. T., Baumeister, R. F., Dewall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., Brewer, L. E., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, *92*(2), 325–336. https://doi.org/10.1037/0022-3514.92.2.325

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: “Anterior insular cortex mediates bodily sensibility and social anxiety” by Terasawa et al.,(2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234. https://doi.org/10.1093/scan/nss140

Garfinkel, S. N., & Critchley, H. D. (2016). Threat and the Body: How the Heart Supports Fear Processing. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.10.005

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 65–74. https://doi.org/10.1016/j.biopsycho.2014.11.004

Giner-Sorolla, R., & Chapman, H. A. (2017). Beyond Purity: Moral Disgust Toward Bad Character. *Psychological Science*, *28*(1), 80–91. https://doi.org/10.1177/0956797616673193

Glick, G., & Braunwald, E. (1965). Relative Roles of the Sympathetic and Parasympathetic Nervous Systems. *Circulation Research*, *16*, 363–375. https://doi.org/10.1161/01.RES.16.4.363

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371. https://doi.org/10.1016/j.cognition.2009.02.001

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105 LP – 2108. http://science.sciencemag.org/content/293/5537/2105.abstract

Grinberg, M., & Hristova, E. (2013). *Moral Judgments : Studying People with Helping Professions*. 704–709.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291. https://doi.org/10.1017/S0048577201393198

Gu, J., Zhong, C. B., & Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *Journal of Experimental Psychology: General*, *142*(2), 307–312. https://doi.org/10.1037/a0029549

Gu, X., & FitzGerald, T. H. B. (2014). Interoceptive inference: Homeostasis and decision-

making. *Trends in Cognitive Sciences*, *18*(6), 260–270. https://doi.org/10.1016/j.tics.2014.02.001

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Halawi, H., Camilleri, M., Acosta, A., Vazquez-Roque, M., Oduyebo, I., Burton, D., Busciglio, I., & Zinsmeister, A. R. (2017). Relationship of gastric emptying or accommodation with satiation, satiety, and postprandial symptoms in health. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *313*(5), G442–G447. https://doi.org/10.1152/ajpgi.00190.2017

Hansson, S. O. (2009). An Agenda for the Ethics of Risk. In *The Ethics of Technological Risk* (pp. 11–23).

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2013). Emotional regulation and bodily sensation: Interoceptive awareness is intact in borderline personality disorder. *Journal of Personality Disorders*, *27*(4), 506–518. https://doi.org/10.1521/pedi_2012_26_049

Herbert, B. M., Herbert, C., Pollatos, O., Weimer, K., Enck, P., Sauer, H., & Zipfel, S. (2012). Effects of short-term food deprivation on interoceptive awareness, feelings and autonomic cardiac activity. *Biological Psychology*, *89*(1), 71–79. https://doi.org/10.1016/j.biopsycho.2011.09.004

Herbert, B. M., Muth, E. R., Pollatos, O., & Herbert, C. (2012). Interoception across modalities: On the relationship between cardiac awareness and the sensitivity for gastric functions. *PLoS ONE*, *7*(5), 1–9. https://doi.org/10.1371/journal.pone.0036646

Herbert, B. M., Ulbrich, P., & Schandry, R. (2007). Interoceptive sensitivity and physical effort: Implications for the self-control of physical load in everyday life. *Psychophysiology*, *44*(2), 194–202. https://doi.org/10.1111/j.1469-8986.2007.00493.x

Higgins, E. T. (1987). Self-Discrepancy: A Theory Relating Self and Affect. *Psychological Review*, *94*(3), 319–340. https://doi.org/10.1080/13549839.2015.1016902

Isomura, T., & Watanabe, K. (2020). Direct gaze enhances interoceptive accuracy. *Cognition*,

195(November 2019), 104113. https://doi.org/10.1016/j.cognition.2019.104113

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jones, M. P., Hoffman, S., Shah, D., Patel, K., & Ebert, C. C. (2003). The water load test: observations from healthy controls and patients with functional dyspepsia. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *284*(6), G896–G904. https://doi.org/10.1152/ajpgi.00361.2002

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *The American Psychologist*, *58*(9), 697—720. https://doi.org/10.1037/0003-066x.58.9.697

Kandasamy, N., Garfinkel, S. N., Page, L., Hardy, B., Critchley, H. D., Gurnell, M., & Coates, J. M. (2016). Interoceptive Ability Predicts Survival on a London Trading Floor. *Scientific Reports*, *6*, 1–7. https://doi.org/10.1038/srep32986

Kang, Y., Williams, L. E., Clark, M. S., Gray, J. R., & Bargh, J. A. (2011). Physical temperature effects on trust behavior: The role of insula. *Social Cognitive and Affective Neuroscience*, *6*(4), 507–515. https://doi.org/10.1093/scan/nsq077

Kant, I. (2018). Groundwork of the Metaphysics of Morals (1785). *Business Ethics: The Big Picture*, 18.

Katkin, E. S. (1985). Blood, sweat, and tears: Individual differences in autonomic self-perception. *Psychophysiology*, *22*(2), 125–137.

Kever, A., Pollatos, O., Vermeulen, N., & Grynberg, D. (2015). Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. *Personality and Individual Differences*, *87*. https://doi.org/10.1016/j.paid.2015.07.014

Khalsa, S. S., Rudrauf, D., Sandesara, C., Olshansky, B., & Tranel, D. (2009). Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. *International Journal of Psychophysiology*, *72*(1), 34–45. https://doi.org/10.1016/j.ijpsycho.2008.08.010

Khalsa, Sahib S., Rudrauf, D., Damasio, A. R., Davidson, R. J., Lutz, A., & Tranel, D. (2008). Interoceptive awareness in experienced meditators. *Psychophysiology*, *45*(4), 671–677. https://doi.org/10.1111/j.1469-8986.2008.00666.x

Khalsa, Sahib S, Rudrauf, D., & Tranel, D. (2009). Interoceptive awareness declines with age. *Psychophysiology*, *46*(6), 1130–1136. https://doi.org/10.1111/j.1469-8986.2009.00859.x

Klabunde, R. (2011). *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins.

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology*, *52*(11), 1432–1440. https://doi.org/10.1111/psyp.12503

Knoll, J. F., & Hodapp, V. (1992). A Comparison between Two Methods for Assessing Heartbeat Perception.pdf. *Psychophysiology*, *29*(2), 218–222. https://doi.org/10.1080/17470218.2016.1181768

Koch, K. L., Hong, S.-P., & Xu, L. (2000). Reproducibility of Gastric Myoelectrical Activity and the Water Load Test in Patients with Dysmotility-like Dyspepsia Symptoms and in Control Subjects. *Journal of Clinical Gastronenterology*, *31*(2), 125–129.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.4324/9781315853161

Krautwurst, S., Gerlach, A. L., Gomille, L., Hiller, W., & Witthöft, M. (2014). Health anxiety - An indicator of higher interoceptive sensitivity? *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(2), 303–309. https://doi.org/10.1016/j.jbtep.2014.02.001

Lacey, B. C., & Lacey, J. I. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. *American Psychologist*, *33*(2), 99–113. https://doi.org/10.1037/0003-066X.33.2.99

Lang, P. J. (1994). The varieties of emotional experience: A meditation on James-Lange theory. Special Issue: The centennial issue of the Psychological Review. *Psychological Review*, *101*(2), 211–221.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at Pictures: Affective, facial, visceral and behavioural reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1080/0449010X.1992.10705887

Lazar, S. W., Kerr, C. E., Wasserman, R. H., Gray, J. R., Greve, D. N., Treadway, M. T., McGarvey, M., Quinn, B. T., Dusek, J. A., Benson, H., Rauch, S. l., Moore, C. I., &

Fischl, B. (2005). Meditation experience is associated with increased cortical thickness Sara. *Neuroreport*, *16*(17), 1893–1897. https://doi.org/10.2214/ajr.174.2.1740487

Lazarus, R.S., & Folkman, S. (1984). Coping and adaptation. *The Handbook of Behavioural Medicine.*, *282325*.

Leopold, C., & Schandry, R. (2001). The heartbeat-evoked brain potential in patients suffering from diabetic neuropathy and in healthy control persons. *Clinical Neurophysiology*, *112*(4), 674–682. https://doi.org/10.1016/S1388-2457(01)00480-1

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. *Frontiers in Human Neuroscience*, *8*(AUG), 1–10. https://doi.org/10.3389/fnhum.2014.00643

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, *33*(2), 273–286. https://doi.org/10.1111/j.1551-6709.2009.01013.x

MacCormack, J. (2016). Feeling "Hangry"? When hunger is conceptualized as an emotion. *Emotion*, *19*(2), 301.

Ma-Kellams, C. (2014). Cross-cultural differences in somatic awareness and interoceptive accuracy: a review of the literature and directions for future research. *Frontiers in psychology*, *5*(1379), 1-9.

Maister, L., & Tsakiris, M. (2014). My face, my heart: Cultural differences in integrated bodily self-awareness. *Cognitive Neuroscience*, *5*(1), 10–16. https://doi.org/10.1080/17588928.2013.808613

Mandler, G., Mandler, J., & Uviller, M. (1958). *Autonomic Feedback : the Perception of Autonomic activity*. *56*(3), 367–373.

Marshall, A. C., Gentsch, A., & Schütz-Bosbach, S. (2018). The interaction between interoceptive and action states within a framework of predictive coding. *Frontiers in Psychology*, *9*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2018.00180

Mckiernan, F., Houchins, J. A., & Mattes, R. D. (2008). Relationships between human thirst, hunger, drinking, and feeding. *Physiology and Behavior*, *94*(5), 700–708.

Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PLoS ONE*, *13*(12), 1–

12. https://doi.org/10.1371/journal.pone.0208034

Mehling, W. E., Daubenmier, J., Price, C. J., Acree, M., Bartmess, E., & Stewart, A. L. (2013). Self-reported interoceptive awareness in primary care patients with past or current low back pain. *Journal of Pain Research*, *6*, 403–418. https://doi.org/10.2147/JPR.S42418

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The Multidimensional Assessment of Interoceptive Awareness (MAIA). *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048230

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the Unexpected: Physiological Responses During Social Interactions With Expectancy-Violating Partners. *Journal of Personality and Social Psychology*, *92*(4), 698–716. https://doi.org/10.1037/0022-3514.92.4.698

Meyerholz, L., Irzinger, J., Witthöft, M., Gerlach, A. L., & Pohl, A. (2019). Contingent biofeedback outperforms other methods to enhance the accuracy of cardiac interoception: A comparison of short interventions. *Journal of Behavior Therapy and Experimental Psychiatry*, *63*(November 2018), 12–20. https://doi.org/10.1016/j.jbtep.2018.12.002

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152. https://doi.org/10.1016/j.tics.2006.12.007

Mikulas, W. L. (2011). Mindfulness: Significant Common Confusions. *Mindfulness*, *2*(1), 1–7. https://doi.org/10.1007/s12671-010-0036-z

Mill, J. S. (1998). *Utilitarianism* (R. Crisp,). Oxford University Press.

Miller, L.C., Murphy, R. and Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology*, *41*(2), 397.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion (Washington, D.C.)*, *14*(3), 573—587. https://doi.org/10.1037/a0035361

Montano, N., Tobaldini, E., & Porta, A. (2012). The Autonomic Nervous System. In A. Chouker (Ed.), *Stress Challenges and Immunity in Space: From Mechanisms to Monitoring and Preventive Strategies* (pp. 71–86). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22272-6_6

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who Shalt Not Kill ? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557.

Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–1899. https://doi.org/10.1162/jocn.2009.21367

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., & Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences*, *105*(35), 13169 LP – 13173. http://www.pnas.org/content/105/35/13169.abstract

Murphy, J, Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2018). Validation of the interoceptive accuracy scale (IAS) supports distinction between self-reported interoceptive accuracy and awareness. doi:10.31234/osf.io/fqgu4

Nahab, F. B., Kundu, P., Gallea, C., Kakareka, J., Pursley, R., Pohida, T., Miletta, N., Friedman, J., & Hallett, M. (2011). The Neural Processes Underlying Self-Agency. *Cerebral Cortex*, *21*(1), 48–55. https://doi.org/10.1093/cercor/bhq059

Naqvi, Nasir, H., & Bechara, A. (2010). The Insula and Drug Addiction: an interoceptive view of pleasure, urges, and decision-making. *Brain Structure and Function*, *214*(0), 435–450. https://doi.org/10.1007/s00429-010-0268-7.The

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: emotion and action in a simulated three-dimensional "trolley problem". *Emotion*, *12*(2), 364–370. https://doi.org/10.1037/a0025561

Nentjes, L., Meijer, E., Bernstein, D., Arntz, A., & Medendorp, W. (2013). Brief Communication: Investigating the Relationship Between Psychopathy and Interoceptive Awareness. *Journal of Personality Disorders*, *27*(5), 617–624. https://doi.org/10.1521/pedi_2013_27_105

Nyholm, S., & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? *Ethical Theory and Moral Practice*, *19*(5), 1275–1289. https://doi.org/10.1007/s10677-016-9745-2

Pan, X., & Slater, M. (2011). Confronting a Moral Dilemma in Virtual Reality: A Pilot Study.

Proceedings of the 25th BCS Conference on Human-Computer Interaction SE - BCS-HCI '11, 46–51. citeulike-article-id:13326879%5Cnhttp://portal.acm.org/citation.cfm?id=2305326

Park, G., Kappes, A., Rho, Y., & Van Bavel, J. J. (2016). At the heart of morality lies neuro-visceral integration: Lower cardiac vagal tone predicts utilitarian moral judgment. *Social Cognitive and Affective Neuroscience*, *11*(10), 1588–1596. https://doi.org/10.1093/scan/nsw077

Parkin, L., Morgan, R., Rosselli, A., Howard, M., Sheppard, A., Evans, D., Hawkins, A., Martinelli, M., Golden, A. M., Dalgleish, T., & Dunn, B. (2014). Exploring the Relationship Between Mindfulness and Cardiac Perception. *Mindfulness*, *5*(3), 298–313. https://doi.org/10.1007/s12671-012-0181-7

Parton, D. M., & McGinley, J. J. (2019). At the heart of harm: Cardiac substrates of action-based aversion to harm. *Personality and Individual Differences*, *151*(July 2018), 109497. https://doi.org/10.1016/j.paid.2019.07.007

Patil, I. (2015). Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion. *Journal of Cognitive Psychology*, *27*(3), 349-366.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2020). Reasoning Supports Utilitarian Resolutions to Moral Dilemmas Across Diverse Measures. *Journal of Personality and Social Psychology*, 1–42. https://doi.org/10.1037/pspp0000281

Pinna, T., & Edwards, D. J. (2020). A Systematic Review of Associations Between Interoception, Vagal Tone, and Emotional Regulation: Potential Applications for Mental Health, Wellbeing, Psychological Flexibility, and Chronic Conditions. *Frontiers in Psychology*, *11*(August), 1–15. https://doi.org/10.3389/fpsyg.2020.01792

Pollatos, O., Füstös, J., & Critchley, H. D. (2012). On the generalised embodiment of pain:

How interoceptive sensitivity modulates cutaneous pain perception. *Pain*, *153*(8), 1680–1686. https://doi.org/10.1016/j.pain.2012.04.030

Pollatos, O., Herbert, B. M., Kaufmann, C., Auer, D. P., & Schandry, R. (2007a). Interoceptive awareness, anxiety and cardiovascular reactivity to isometric exercise. *International Journal of Psychophysiology*, *65*(2), 167–173. https://doi.org/10.1016/j.ijpsycho.2007.03.005

Pollatos, O., Herbert, B. M., Matthias, E., & Schandry, R. (2007b). Heart rate response after emotional picture presentation is modulated by interoceptive awareness. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2006.09.003

Pollatos, O., Kirsch, W., & Schandry, R. (2005). On the relationship between interoceptive awareness, emotional experience, and brain processes. *Cognitive Brain Research*, *25*(3), 948–962. https://doi.org/10.1016/j.cogbrainres.2005.09.019

Porges, S. (1993). Body Perception Questionnaire. *Laboratory of Developmental Assessment, University of Maryland*.

Pramme, L., Larra, M. F., Schächinger, H., & Frings, C. (2016). Cardiac cycle time effects on selection efficiency in vision. *Psychophysiology*, *53*(11), 1702–1711. https://doi.org/10.1111/psyp.12728

Prout, B. J. (1967). Independence of the galvanic skin reflex from the vasoconstrictor reflex in man. *Journal of Neurology, Neurosurgery, and Psychiatry*, *30*(4), 319–324. https://doi.org/10.1136/jnnp.30.4.319

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Rani, N. J., & Rao, P. V. K. (1994). Body Awareness and Yoga Training. *Perceptual and Motor Skills*, *79*(3), 1103–1106. https://doi.org/10.2466/pms.1994.79.3.1103

Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion*, *18*(7), 1009–1023. https://doi.org/10.1037/emo0000413

Ring, C., & Brener, J. (2018). Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. *Psychophysiology*, 1–10. https://doi.org/10.1111/psyp.13084

Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: A cautionary tale about interoceptive awareness. *Biological Psychology*, *104*, 193–198. https://doi.org/10.1016/J.BIOPSYCHO.2014.12.010

Riva, G., Baños, R. M., Botella, C., Mantovani, F., & Gaggioli, A. (2016). Transforming experience: The potential of augmented reality and virtual reality for enhancing personal and clinical change. In *Frontiers in Psychiatry* (Vol. 7, Issue SEP). https://doi.org/10.3389/fpsyt.2016.00164

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*(2), 165–184. https://doi.org/10.1023/A:1019923923537

Rozin, P. (2015). Disgust, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Second Edi, Vol. 6). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26020-0

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality & Social Psychology*, *76*(5), 805.

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology and Behavior*, *130*, 127–134. https://doi.org/10.1016/j.physbeh.2014.04.002

Schaefer, M., Egloff, B., Gerlach, A. L., & Witthöft, M. (2014). Improving heartbeat perception in patients with medically unexplained symptoms reduces symptom distress. *Biological Psychology*, *101*(1). https://doi.org/10.1016/j.biopsycho.2014.05.012

Schandry, R. (1981). Heart Beat Perception and Emotional Experience. *Psychophysiology*, *18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality & Social Psychology Bulletin*, *34*(8), 1096–1109. https://doi.org/10.1177/0146167208317771

Schneider, T., Ring, C., & Katkin, E. (1998). A test of the validity of the method of constant stimuli as an index of heartbeat detection. *Psychophysiology*, *35*(1), 86–89.

Schoen, E. J. (1957). Minimum urine total solute concentration in response to water loading in normal men. *Journal of Applied Physiology*, *35*(1), 86–89.

Schulz, A.. (2015). Interoception. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, *12*, 614–620. https://doi.org/10.1016/B978-0-08-097086-8.14145-5

Schulz, A., Matthey, J. H., Vögele, C., Schaan, V., Schächinger, H., Adler, J., Beutel, M. E., & Michal, M. (2016). Cardiac modulation of startle is altered in depersonalization-/derealization disorder: Evidence for impaired brainstem representation of baro-afferent neural traffic. *Psychiatry Research*, *240*, 4–10. https://doi.org/10.1016/j.psychres.2016.03.051

Schwitzgebel, E., & Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, *27*(2), 135–153. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0017.2012.01438.x/full

Seery, M. D. (2011). Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience and Biobehavioral Reviews*, *35*(7), 1603–1610. https://doi.org/10.1016/j.neubiorev.2011.03.003

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., & Critchley, H. D. (2013). Extending predictive processing to the body: Emotion as interoceptive inference. *Behavioral and Brain Sciences*, *36*(3), 227–228. https://doi.org/10.1017/S0140525X12002270

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). https://doi.org/10.1098/rstb.2016.0007

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *3*(JAN), 1–16. https://doi.org/10.3389/fpsyg.2011.00395

Sherwood, A., Dolan, C. A., & Light, K. C. (1990). Hemodynamics of Blood Pressure Responses During Active and Passive Coping. *Psychophysiology*, *27*(6), 656–668. https://doi.org/10.1111/j.1469-8986.1990.tb03189.x

Silvia, P. J., Eddington, K. M., Beaty, R. E., Nusbaum, E. C., & Kwapil, T. R. (2013). Gritty people try harder: Grit and effort-related cardiac autonomic activity during an active coping challenge. *International Journal of Psychophysiology*, *88*(2), 200–205. https://doi.org/10.1016/j.ijpsycho.2013.04.007

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–340. https://doi.org/10.1016/j.tics.2009.05.001

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI*, *3*(December), 1–47. https://doi.org/10.3389/frobt.2016.00074

Sterling, P. (2015). Principles of allostasis: Optimal design, predictive regulation, pathophysiology, and rational therapeutics. *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*, 17–64. https://doi.org/10.1017/CBO9781316257081.004

Stevenson, R. J., Mahmut, M., & Rooney, K. (2015). Individual differences in the interoceptive states of hunger, fullness and thirst. *Appetite*, *95*, 44–57. https://doi.org/10.1016/j.appet.2015.06.008

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458. https://doi.org/10.1016/j.cognition.2011.01.018

Suzuki, A., Hirota, A., Takasawa, N., & Shigemasu, K. (2003). Application of the somatic marker hypothesis to individual differences in decision making. *Biological Psychology*, *65*(1), 81–88. https://doi.org/10.1016/S0301-0511(03)00093-0

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, *51*(13), 2909–2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014

Tabri, N., Hollingshead, S. J., & Wohl, M. J. A. (2020). *Framing COVID-19 as an Existential Threat Predicts Anxious Arousal and Prejudice towards Chinese People Like.* Preprints. https://doi.org/10.1017/CBO9781107415324.004

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are we more moral than we think? Exploring the role of affect in moral behavior and moral forecasting. *Psychological Science*, 22(4), 553-558.

Teper, R., Zhong, C., & Inzlicht, M. (2015). How Emotions Shape Moral Behavior : Some Answers (and Questions) for the Field of Moral Psychology. *Social and Personality Psychology Compass*, *9*(1), 1–14.

Terasawa, Y., Shibata, M., Moriguchi, Y., & Umeda, S. (2013). Anterior insular cortex mediates bodily sensibility and social anxiety. *Social Cognitive and Affective Neuroscience*, *8*(3), 259–266. https://doi.org/10.1093/scan/nss108

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, Physiological, and Behavioral Effects of Threat and Challenge Appraisal. *Journal of Personality and Social Psychology*, *65*(2), 248–260. https://doi.org/10.1037/0022-3514.65.2.248

Tracy, J. L., Steckler, C. M., & Heltzel, G. (2019). The physiological basis of psychological disgust and moral judgments. *Journal of Personality and Social Psychology*, *116*(1), 15–32. https://doi.org/10.1037/pspa0000141

Tsakiris, M. (2020). How should the political animals of the 21st century feel?: Comment on "The sense of should: A biologically-based framework for modelling social pressure" by JE Theriault et al.,*Physics of Life Reviews*, *January*, 19–22.

Tsakiris, M., & Critchley, H. (2016). Interoception beyond homeostasis: affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708). http://rstb.royalsocietypublishing.org/content/371/1708/20160002.abstract

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure & Function*, *214*(5–6), 629–643. https://doi.org/10.1007/s00429-010-0261-1

Van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P. C., Schulz, A., & Herbert, B. M. (2016). The water load test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PLoS ONE*, *11*(9), 1–14. https://doi.org/10.1371/journal.pone.0163574

Veeraraghavan, H., Bird, N., Atev, S., & Papanikolopoulos, N. (2007). Classifiers for driver

activity monitoring. *Transportation Research Part C: Emerging Technologies*, *15*(1), 51–67. https://doi.org/10.1016/j.trc.2007.01.001

Verdejo-García, A., & Bechara, A. (2009). A somatic marker theory of addiction. *Neuropharmacology*, *56*(SUPPL. 1), 48–62. https://doi.org/10.1016/j.neuropharm.2008.07.035

Vicario, C. M., Kuran, K. A., Rogers, R., & Rafal, R. D. (2018). The effect of hunger and satiety in the judgment of ethical violations. *Brain and Cognition*, *125*, 32–36. https://doi.org/https://doi.org/10.1016/j.bandc.2018.05.003

von Leupoldt, A., Keil, A., Chan, P.-Y. S., Bradley, M. M., Lang, P. J., & Davenport, P. W. (2010). Cortical sources of the respiratory-related evoked potential. *Respiratory Physiology & Neurobiology*, *170*(2), 198–201. https://doi.org/10.1016/J.RESP.2009.12.006

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*(3), 247–253. https://doi.org/10.1111/j.1467-9280.2007.01884.x

Werner, N. S., Duschek, S., Mattern, M., & Schandry, R. (2009). The relationship between pain perception and interoception. *Journal of Psychophysiology*, *23*(1), 35–42. https://doi.org/10.1027/0269-8803.23.1.35

Werner, N. S., Jung, K., Duschek, S., & Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology*, *46*(6), 1123–1129. https://doi.org/10.1111/j.1469-8986.2009.00855.x

Werner, N. S., Peres, I., Duschek, S., & Schandry, R. (2010). Implicit memory for emotional words is modulated by cardiac perception. *Biological Psychology*, *85*(3), 370–376. https://doi.org/10.1016/j.biopsycho.2010.08.008

Whitehead, W.E., Drescher, V. M., & Heiman, P. (1977). Relation of heart rate control to heartbeat perception. *Biofeedback and Self-Regulation*, *2*(4), 317–392.

Whitehead, William E., & Drescher, V. M. (1980). Perception of Gastric Contractions and Self-Control of Gastric Motility. *Psychophysiology*, *17*(6), 552–558. https://doi.org/10.1111/j.1469-8986.1980.tb02296.x

Wiens, S., Mezzacappa, E. S., & Katkin, E. S. (2000). Heartbeat detection and the experience of emotions. *Cognition and Emotion*, *14*(3), 417–427.

https://doi.org/10.1080/026999300378905

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*(5901), 606–607. https://doi.org/10.1126/science.1162548

Wilson, T., & Gilbert, D. (2003). Affective Forecasting. Knowing What to Want (part 1). *Advances in Experimental Social Psychology*, *35*, 345–411.

Wright, P., He, G., Shapira, N. A., Goodman, W. K., & Liu, Y. (2004). Disgust and the insula: fMRI responses to pictures of mutilation and contamination. *NeuroReport*, *15*(15), 2347–2351. https://doi.org/10.1097/00001756-200410250-00009

Yoder, K. J., & Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia*, *60*(1), 39–45. https://doi.org/10.1016/j.neuropsychologia.2014.05.022

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to Ventromedial Prefrontal Cortex Impairs Judgment of Harmful Intent. *Neuron*, *65*(6), 845–851. https://doi.org/10.1016/j.neuron.2010.03.003

Zaki, J., Davis, J. I., & Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage*, *62*(1), 493–499. https://doi.org/10.1016/j.neuroimage.2012.05.012

Zeng, J., Zuo, X.-L., Li, Y.-Q., Wei, W., & Lv, G.-P. (2007). Tegaserod for dyspepsia and reflux symptoms in patients with chronic constipation: an exploratory open-label study. *European Journal of Clinical Pharmacology*, *63*(6), 529–536. https://doi.org/10.1007/s00228-007-0287-3

Zhang, L., Kong, M., & Li, Z. (2017). Emotion regulation difficulties and moral judgment in different domains: The mediation of emotional valence and arousal. *Personality and Individual Differences*, *109*, 56–60. https://doi.org/10.1016/j.paid.2016.12.049